Computers & Security 144 (2024) 103964

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

From COBIT to ISO 42001: Evaluating cybersecurity frameworks for opportunities, risks, and regulatory compliance in commercializing large language models

Timothy R. McIntosh [a,b,*], Teo Susnjak [c], Tong Liu [c], Paul Watters [d], Dan Xu [e], Dongwei Liu [f], Raza Nowrozy [g], Malka N. Halgamuge [h]

[a] La Trobe University, Bundoora, VIC, Australia
[b] Cyberoo Pty Ltd, Surrey Hills, NSW, Australia
[c] Massey University, Auckland, New Zealand
[d] Cyberstronomy Pty Ltd, Ballarat, VIC, Australia
[e] ANZ Bank, Melbourne, VIC, Australia
[f] Coles Group, Hawthorn East, VIC, Australia
[g] Untapped Holdings, Melbourne, VIC, Australia
[h] RMIT University, Melbourne, VIC, Australia

ARTICLE INFO

Keywords:
Cybersecurity frameworks
Large language models
Risk management
AI governance
Cyber resilience
Information security

ABSTRACT

This study investigated the integration readiness of four predominant cybersecurity *Governance, Risk and Compliance* (GRC) frameworks – NIST CSF 2.0, COBIT 2019, ISO 27001:2022, and the latest ISO 42001:2023 – for the opportunities, risks, and regulatory compliance when adopting *Large Language Models* (LLMs), using qualitative content analysis and expert validation. Our analysis, with both LLMs and human experts in the loop, uncovered potential for LLM integration together with inadequacies in LLM risk oversight of those frameworks. Comparative gap analysis has highlighted that the new ISO 42001:2023, specifically designed for *Artificial Intelligence* (AI) management systems, provided most comprehensive facilitation for LLM opportunities, whereas COBIT 2019 aligned most closely with the European Union AI Act. Nonetheless, our findings suggested that all evaluated frameworks would benefit from enhancements to more effectively and more comprehensively address the multifaceted risks associated with LLMs, indicating a critical and time-sensitive need for their continuous evolution. We propose integrating human-expert-in-the-loop validation processes as crucial for enhancing cybersecurity frameworks to support secure and compliant LLM integration, and discuss implications for the continuous evolution of cybersecurity GRC frameworks to support the secure integration of LLMs.

1. Introduction

Cybersecurity frameworks, such as the *National Institute of Standards and Technology* (NIST) *Cybersecurity Framework* (CSF), *Control Objectives for Information and Related Technology* (COBIT), and *International Organization for Standardization* (ISO) 27001, are indispensable templates for diverse organizational sectors, with their dominance supported by recent industry reports and surveys (Kure et al., 2022; Sulistyowati et al., 2020; Tissir et al., 2021). The NIST CSF, for instance, garners substantial endorsements both domestically and abroad, highlighted by its extensive academic citations (Dedeke, 2017; Tissir et al., 2021). Similarly, ISO 27001 boasts over 40,000 global certifications, emphasizing its stature in information security management (Hsu et al., 2016; Mirtsch et al., 2020). COBIT 2019, validated by several surveys, continues to be a premier choice for IT governance professionals (Atrinawati et al., 2021; Febriyani et al., 2022; Nugraheni et al., 2022). Recently introduced in December 2023, the ISO 42001:2023 sets forth requirements for establishing and improving an Artificial Intelligence (AI) Management System, yet there has been no systematic academic analysis of its advantages and disadvantages due to its novelty. Incorporating structured cyber risk navigation and enabling organizations to create custom cybersecurity governance tools, these frameworks continually evolve through a feedback-based approach of frequent revisions and ongoing collaborations with both public and private sectors, thereby addressing the dynamic technological landscape and emerging threats (Kure et al., 2022; Sulistyowati et al., 2020; Tissir et al., 2021; Yusif and Hafeez-Baig, 2021). However, it should be noted

* Corresponding author at: La Trobe University, Bundoora, VIC, Australia.
E-mail address: t.mcintosh@latrobe.edu.au (T.R. McIntosh).

https://doi.org/10.1016/j.cose.2024.103964
Received 10 November 2023; Received in revised form 4 June 2024; Accepted 18 June 2024
Available online 22 June 2024

that, apart from ISO 42001:2023, none of these frameworks specifically cover AI adaptation and governance. Operationally guiding a range of functions that span asset categorization, control selection, training, and audits, these frameworks not only set the industry benchmark and serve as critical tools that consultants frequently leverage to evaluate the robustness of their clients' cybersecurity strategies, but also act as foundational pillars for academics, thereby significantly contributing to the enhancement of cybersecurity workforce expertise (Bozkus Kahyaoglu and Caliyurt, 2018; Kure et al., 2022; Yeoh et al., 2022). However, these frameworks face challenges in adapting their comprehensive controls to specific organizational environments, resources, and risk appetites. The integration of newer technologies, such as cloud computing and *Artificial Intelligence* (AI), adds to the complexity, necessitating nuanced revisions to these frameworks (Kabanda et al., 2018; Radanliev et al., 2018; Tawalbeh et al., 2020; Tissir et al., 2021; Tvaronavičienė et al., 2020). The cybersecurity landscape is slowly being transformed by the incorporation of *Large Language Models* (LLMs), a change that Deloitte[1] and KPMG[2] have already embraced, enhancing cybersecurity audit and operation capabilities and offering new opportunities for innovation in policy and compliance (Darraj et al., 2019; McIntosh et al., 2023a; Yang et al., 2023). With over 92% of Fortune 500 companies utilizing the OpenAI platform,[3] LLMs are influencing corporate practices across various sectors, not just within cybersecurity. Nonetheless, LLMs can introduce their own set of challenges, particularly the risk of generating unreliable or 'hallucinated' content, complicating their integration into existing cybersecurity measures (Ji et al., 2023; Kaur et al., 2023; McIntosh et al., 2023a,b, 2024a).

The ongoing discourse within the research community has been critical of the practicality and scientific underpinning of prevailing cybersecurity frameworks. Critics have pointed out the lack of empirical evidence to substantiate the effectiveness of these frameworks in enhancing security outcomes (Bayuk, 2013; Katina and Keskin, 2021; Malaivongs et al., 2022; Manuel et al., 2022; Paskauskas, 2022). They highlight the frameworks' propensity for detailed taxonomies of controls over actionable guidance for organizational-specific risk profiles (Argyridou et al., 2023; Bozkus Kahyaoglu and Caliyurt, 2018). Further, it is argued that the frameworks do not sufficiently account for the multi-disciplinary nature required to address complex socio-technical challenges, often presenting a limited technical compliance viewpoint (Bayuk, 2013; Ekambaranathan et al., 2023; Shim et al., 2020). Notably, existing frameworks have been found lacking in their coverage of new technologies, such as cloud computing and IoT (Darraj et al., 2019; Kaur et al., 2023). Conversely, supporters of these frameworks suggest they play a crucial role in raising awareness, unifying industry language, and embodying agreed-upon best practices, despite not ensuring security (Cho et al., 2015; Hajny et al., 2021; Manuel et al., 2022). Acknowledging these limitations, there is a concerted effort in recent research to augment cybersecurity frameworks by incorporating insights from the domains of security economics (Ekelund and Iskoujina, 2019; Radanliev et al., 2018; Rathod and Hämäläinen, 2017), behavioral psychology (King et al., 2018; Maalem Lahcen et al., 2020), and system safety (Li and Liu, 2021; Taherdoost, 2022). The introduction of generative AI, such as LLMs, has fueled further debate, to include how to evolve these frameworks to safely utilize AI for security automation while managing associated risks, like model hallucinations (McIntosh et al., 2023a, 2024b). This synthesis of perspectives implies that while cybersecurity frameworks are beneficial starting points, they necessitate contextual adjustments and enhancements to drive substantive improvements in security programs.

This study assessed the readiness of leading cybersecurity frameworks (*i.e.*, COBIT 2019, ISO 27001:2022 (general purpose GRC), ISO 42001:2023 (AI Management System, or AIMS), and the NIST CSF 2.0) — chosen for their comprehensive coverage of *Governance, Risk, and Compliance* (GRC) principles and their widespread adoption as one-stop shops for organizational GRC blueprints — in addressing the challenges and leveraging the opportunities presented by the integration of LLMs into cybersecurity operations. Amid a landscape where diverse bodies vie to set industry standards with frameworks that differ greatly in coverage, emphasis, and levels of abstraction, these were chosen for their robust encapsulation of GRC principles. Furthermore, the inherent abstractness and principle-based approach of these frameworks lend a degree of subjectivity to their interpretation, paralleling the diverse legal interpretations encountered in legislation. To address this, our study innovatively employs both LLMs and human experts in a loop, fostering a consensus-based interpretation to minimize disagreements. The study's motivation arose from NIST's formal release of NIST CSF 2.0,[4] and the enactment of the *European Union* (EU) AI Act.[5] Our analysis focused on the secure, ethical use of generative AI, with an emphasis on LLMs, examining COBIT 2019, ISO 27001, ISO 42001, and NIST CSF 2.0, with equal rigor to deliver a comprehensive evaluation of their efficacy in the GRC context. Our research identified critical deficiencies in these frameworks, notably in the areas of human oversight, validation controls, and adherence to compliance—a crucial consideration in light of technologies like LLMs. Our comprehensive evaluation covered: (1) assessing the frameworks' support for opportunities of adopting and integrating LLMs, (2) evaluating the inclusion of provisions for LLM risk mitigation, and human oversight and validation, and (3) determining the preparedness of the frameworks to align with the EU AI Act's main provisions, set to regulate the rapidly advancing generative AI industry that brought LLMs to global prominence in 2023. The intent of this research was not to directly instruct organizations on integrating LLMs into their GRC practices, but to stimulate informed discourse for timely enhancements to GRC frameworks. Such updates would bolster organizational reliance on these frameworks as they assimilate LLM technologies. Our findings have signaled an urgent need for framework modernization to address risks and compliance issues associated with emergent AI technologies, while capitalizing on the opportunities of their adoption and integration, through improved regulatory compliance and secure LLM guidelines. The analysis intends to ignite a vital, evidence-driven debate on the necessity for regular updates to cybersecurity standards, in the face of rapid technological evolution like LLMs.

This research makes several key contributions to the intersection of cybersecurity frameworks and LLM governance:

(1) We have provided one of the first academic evaluations of the preparedness of leading cybersecurity frameworks for integrating LLMs, revealing gaps in risk oversight.
(2) Our analysis of integration potential versus risk provisions has highlighted the need for a multi-dimensional approach as frameworks evolve to support LLMs.
(3) We have identified a lack of controls for managing LLM hallucination risks across frameworks, illuminating an issue that likely extends beyond the analyzed standards.
(4) Our findings have revealed the urgency of continuous evolution and timely version adoption for frameworks to address emerging technologies like LLMs, and in anticipation of regulatory shifts such as the forthcoming EU AI Act.

[1] https://www2.deloitte.com/uk/en/pages/deloitte-analytics/articles/embedding-controls-and-risk-mitigations-throughout-the-generative-ai-development-lifecycle.html.

[2] https://kpmg.com/xx/en/blogs/home/posts/2023/02/all-eyes-on-transforming-the-audit-with-ai.html.

[3] https://www.cnbc.com/2023/11/06/openai-announces-more-powerful-gpt-4-turbo-and-cuts-prices.html.

[4] https://csrc.nist.gov/pubs/cswp/29/the-nist-cybersecurity-framework-csf-20/final.

[5] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

The rest of this study is organized as follows: Section 2 introduces related works. Section 3 explains our study methodology. Section 4 demonstrates the study design and validation. Section 5 presents the results of our analysis. Section 6 discusses the key findings of our study, and explores future research directions. Section 7 concludes this study. Appendix proposes a roadmap to revise those cybersecurity frameworks.

2. Related work

This section examines studies related to the critical evaluation and enhancement of cybersecurity frameworks as well as the emerging threats posed by LLMs in the cybersecurity domain.

2.1. Cybersecurity framework evaluation and improvements

There is a wealth of literature focused on the critical examination of mainstream cybersecurity frameworks, including the NIST CSF, COBIT, and ISO 27001. One recurrent theme centers on the limited empirical evidence supporting the real-world efficacy of these frameworks in improving security outcomes. For instance, a few studies (*e.g.*, Garvey and Patel, 2014; Gourisetti et al., 2020, 2021; Hitchcox, 2020; Malaivongs et al., 2022 and Renaud and Ophoff, 2021) found limited empirical data on cybersecurity framework effectiveness, and called for more rigorous, evidence-based studies on implementation impacts. Others have scrutinized the scientific validity of the risk management models embedded within frameworks. Some studies (*e.g.*, Argyridou et al., 2023; Bozkus Kahyaoglu and Caliyurt, 2018; Hitchcox, 2020; Katina and Keskin, 2021; Kissoon, 2020; Manuel et al., 2022 and Syafrizal et al., 2020) argued that a few main controls and recommendations of many cybersecurity frameworks lacked adequate theoretical and mathematical rigor or implementation practicality. Along similar lines, some studies (*e.g.*, Bayuk, 2013; Gourisetti et al., 2020; Hitchcox, 2020 and Syafrizal et al., 2020) highlighted scientific inconsistencies in some widely used frameworks, potentially suggesting that those frameworks were written based on the limited experiences of their authors or authoring groups.

Beyond scientific rigor, researchers have identified gaps in framework coverage and practical guidance, *e.g.*, deficient or lacking security controls in the NIST CSF for emerging technologies (Darraj et al., 2019; Karie et al., 2021; Kaur et al., 2023), and limited actionable direction on control implementation (Goel et al., 2020). To address these limitations, studies have recommended integrating complementary perspectives into frameworks, including security economics (Ekelund and Iskoujina, 2019; Radanliev et al., 2018; Rathod and Hämäläinen, 2017), behavioral psychology (King et al., 2018; Maalem Lahcen et al., 2020), and system safety fields (Li and Liu, 2021; Taherdoost, 2022).

2.2. LLM threats in cybersecurity

The integration of LLMs into cybersecurity processes has prompted a surge in research focused on identifying and mitigating potential threats posed by these technologies. A critical issue is content hallucination, where LLMs generate plausible but factually incorrect information, which can have serious implications for cybersecurity, particularly in areas such as threat intelligence and incident response, where accuracy is paramount (Ji et al., 2023; McIntosh et al., 2023a). LLMs can create convincing yet entirely fabricated cyber threat reports, potentially leading to misinformed security measures (Gupta et al., 2023; Liu et al., 2023a; Weidinger et al., 2021). LLMs were found to have provided misleading information, which could lead to inadequate or erroneous vulnerability management, potentially leaving systems exposed to unaddressed security risks (Liu et al., 2023b; Szabó and Bilicki, 2023). LLMs can be exploited to produce harmful or toxic outputs used for adversarial attacks, or be prompted to generate content that is seemingly benign but contains subtle manipulations intended to deceive or cause harm (Cheong et al., 2022; Ji et al., 2023; Qi et al., 2023). LLMs can generate discriminatory biases within their outputs, which could inadvertently lead to biased cybersecurity practices, where such biases could manifest in security tools that rely on LLMs, potentially leading to unequal protection measures across different user groups (Burton, 2023; Zhang and Kamel Boulos, 2023).

To navigate these threats, the academic community has advocated for a human-centric approach to LLM governance in cybersecurity, which includes the development of frameworks that prioritize transparency, human oversight, and continuous evaluation of LLM outputs, to ensure that while LLMs can significantly contribute to cybersecurity efforts, they do so in a manner that is secure, ethical, and aligned with the overarching objectives of cyber defense strategies (Asad et al., 2023; Ukil et al., 2023).

3. Methodology

This section outlines our methodology, focusing on the criteria used to select cybersecurity frameworks and the analytical approach adopted for evaluation. Our methodology evaluated their effectiveness against both existing and emerging threats, particularly those introduced by LLMs. We began by examining the core elements of each framework, assessing their efficacy in the current threat landscape. Following this, we identified potential vulnerabilities to LLM threats, before making our final recommendations.

3.1. Theoretical foundation

This research investigated the critical interplay between cybersecurity GRC and the evolving LLM landscape, emphasizing the essential role of understanding different challenges and prospects of LLMs in line with existing and expected regulations, thereby providing a holistic view of LLM integration into cybersecurity strategies.

Governance: Cybersecurity governance refers to the principles and practices designed to safeguard digital assets and data (Katina and Keskin, 2021; Yusif and Hafeez-Baig, 2021). In the scope of this study, governance provides the blueprint to understand the structure and intent of cybersecurity frameworks. The main tenet here is that governance structures should be both robust and agile, especially in the face of novel AI-driven challenges such as LLMs (Asad et al., 2023; Ukil et al., 2023).

Risk Management: The cornerstone of any robust cybersecurity strategy, risk management revolves around identifying, assessing, and addressing vulnerabilities and threats (Jarjoui and Murimi, 2021; McIntosh et al., 2023a). LLMs introduce new avenues of risk: be it through generating malicious content or identifying gaps in security frameworks. Understanding this dynamic ensures a comprehensive evaluation of the frameworks under study.

Compliance and Legislative Aspects: The EU AI Act, anticipated to come into force in May 2025, underscores the critical importance of compliance in the era of advanced AI (Dhirani et al., 2023; Khan and Mer, 2023). We believe the EU AI Act is likely to become the blueprint for other jurisdictions to propose their own AI regulations, akin to how the EU *General Data Protection Regulation* (GDPR) has set the blueprint for others to enhance their Privacy Act. AI solutions, including LLMs, while not yet required to adhere strictly to the Act until May 2025, would benefit from taking its main provisions into consideration early in their development roadmap, to prevent potential last-minute rushes or costly system revisions later on. While we recognize the EU AI Act's attempt to standardize AI governance as a pioneering effort, we acknowledge its evolving nature and the imperative for entities within its jurisdiction to comply, despite the perceived limitations from an organizational perspective.

AI Ethics and Oversight: As LLMs find utility in the research methodology and the EU AI Act gains prominence, ethical facets associated with LLM deployment warrant serious deliberation (Dhirani et al.,

Table 1
Evaluation of cybersecurity frameworks (✓: comprehensive; △: partial; *: selected).

Frameworks	Cybersecurity GRC blueprint suitability			Final selection
	Governance	Risk management	Compliance	
NIST CSF 2.0	✓	✓	✓	*
COBIT 2019	✓	✓	✓	*
ISO 27001:2022	✓	✓	✓	*
ISO 42001:2023	✓	✓	✓	*
CIS controls		✓		
Australian essential 8		✓		
ITIL	△			
GDPR			✓	
HIPAA			✓	
PCI DSS		△	✓	
FAIR	△	✓		
CIS RAM	△	✓		
SABSA	✓	✓		
TOGAF	✓	△	△	
MITRE ATT&CK		✓		
CSA standards	△	△	△	

2023; Floridi et al., 2021; Weidinger et al., 2021). It extends beyond mere deployment, and upholds the principles of transparency in LLM processes, ensuring that LLMs remain accountable for their decisions, and upholding fairness, particularly when those models intersect with or influence cybersecurity protocols (Dhirani et al., 2023; Weidinger et al., 2021). Recognizing and addressing these ethical dimensions can solidify the credibility and trustworthiness of LLMs in cybersecurity contexts.

3.2. Framework selection

All four cybersecurity frameworks considered in this research were sourced directly from their respective official websites, focusing exclusively on those providing comprehensive coverage across governance, risk management, and compliance (GRC), to ensure authenticity, transparency, and accuracy of information, and an integrative approach to cybersecurity that aligns with the multifaceted nature of LLM technologies. In addition, a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis was employed to further evaluate the selected frameworks against the backdrop of LLM technologies, enabling a structured and comprehensive assessment of their capabilities and areas for enhancement.

Drawing our assessment to a close, it became evident from Table 1 that the NIST CSF 2.0, COBIT 2019, ISO 27001:2022 and ISO 42001:2023 emerged as the most suitable choices for the purpose of this study, as the coverage of GRC by all other frameworks are only partial.

To provide clarity on our selection rationale:

- *NIST CSF 2.0*: This framework was selected due to its comprehensive coverage of governance, risk management, and compliance aspects tailored to commercial LLMs, providing a robust structure for cybersecurity practices.
- *COBIT 2019*: This framework was chosen for its strong emphasis on governance and risk management, alongside adaptable compliance measures that can evolve with technological advancements.
- *ISO 27001:2022*: This standard was included for its extensive risk management and compliance protocols, which are adaptable to the needs of commercial LLMs, offering a thorough approach to information security management.
- *ISO 42001:2023*: This framework was selected for its integrative approach to governance, risk management, and compliance, with a focus on continuous improvement and adaptability to new technologies, including LLMs.
- *CIS Controls*: Though it offers vital technical guidelines, it does not cover the broader aspects of governance, risk management, and compliance comprehensively.
- *Australian Essential Eight*: This framework is predominantly tailored to technical standards on Microsoft platforms within Australia, limiting its global applicability.
- *ITIL (Information Technology Infrastructure Library)*: A recognized global standard; however, its primary focus is IT service management, without fully encapsulating the broader spectrum of GRC.
- *PCI DSS (Payment Card Industry Data Security Standard)*: Its primary emphasis is the security assurance of the payment card industry, making its scope more narrow compared to our selected frameworks.
- *FAIR (Factor Analysis of Information Risk), and CIS RAM (Center for Internet Security Risk Assessment Method)*: While both tools emphasize risk assessment, they fall short in providing robust governance structures or clear compliance guidelines, making them less versatile in a GRC-focused approach.
- *SABSA (Sherwood Applied Business Security Architecture)*: While recognized globally, its central thrust is on security architecture, diverging from the comprehensive GRC approach we sought.
- *TOGAF (The Open Group Architecture Framework)*: An enterprise architecture methodology and framework, TOGAF ensures alignment between business and IT, providing strategic governance, efficient risk management, and broad compliance aspects.
- *MITRE ATT&CK (MITRE Corporation Adversarial Tactics, Techniques, and Common Knowledge)*: Primarily known as a knowledge base of attacker behaviors, it is increasingly referenced for cybersecurity risk management and governance. However, its primary focus is not broad GRC, but it provides valuable insights into potential threats and techniques.
- *CSA (Cloud Security Alliance) Standards*: While CSA provides best practices across governance, risk management, and compliance, its focus is primarily on cloud-centric environments, which limits its broader applicability in diverse technological contexts.

3.3. Procedure for LLM-related analysis

To ensure a systematic and rigorous assessment of cybersecurity frameworks concerning LLM integration, our evaluation divided the analysis into defensive and offensive categories:

(1) *Defensive Analysis:*

- *Identification of LLM Integration Potential*: We used the mapping rubric, we assessed each framework for processes amenable to LLM augmentation, correlating opportunities with known LLM capabilities to ensure feasible and beneficial integration.

Table 2
Mapping rubric for framework attributes, LLM characteristics, and EU AI Act readiness.

Cybersecurity framework provisions	LLM capabilities	LLM-related risks[a]	EU AI Act readiness
Process automation (Alromaih et al., 2022; Maglaras et al., 2021)	Natural language understanding and generation (Min et al., 2023; Yang et al., 2023; Zhang et al., 2022)	Misleading content generation (Ji et al., 2023; Min et al., 2023)	Compliance (*Article 13, 52*), transparency (*Article 52-53*), human oversight (*Article 14, 29, 61, 63*).
Real-time analysis (Abie, 2019; Akande et al., 2023; McIntosh et al., 2023a)	Real-time data processing (Min et al., 2023; Zhang et al., 2022)	Bias, unpredictability, latency in responses (Burton, 2023; McIntosh et al., 2023a; Zhang and Kamel Boulos, 2023)	Timely, unbiased AI system responsiveness (*Article 9, 13, 14*). Real-time safeguards against risks (*Article 5, 9, 62, 65*).
Data security and protection (Markopoulou et al., 2019; Sule et al., 2021)	Data analytics (Min et al., 2023; Yang et al., 2023; Zhang et al., 2022), image recognition and generation (Cheong et al., 2022)	Potential for data leakage (Montagna et al., 2023; Winograd, 2023), visual deception, forensic unreliability (Cheong et al., 2022)	Robust protection against data breaches (*Article 15, 70*). Ensuring AI data quality and integrity (*Article 10. Annex IV-2, VII-4.3*).
Continuous monitoring and auditing (Gourisetti et al., 2021; Malatji et al., 2019)	Continuous learning (Min et al., 2023; Zhang et al., 2022)	Over-reliance on outdated training data (Singhal et al., 2023)	Periodic AI assessment (*Article 61, 84*). Addressing outdated data (*Article 10, 43, 61*).
Incident response (Dykstra et al., 2022; McIntosh et al., 2023a)	Automated incident detection and reporting (Gupta et al., 2023; Iturbe et al., 2023)	Misclassification of incidents (Rjoub et al., 2023)	Rapid AI-driven incident recognition (*Article 62, 68*). Mitigate misclassification risks (*Article 10, 13, 14, 15, 43, 61*).
Security awareness training (Khader et al., 2021; Triplett, 2022)	Adaptive training modules (Kasneci et al., 2023)	Distorted reality focus (Ji et al., 2023; Min et al., 2023). Data profiling risks (Weidinger et al., 2021)	AI transparency (*Article 13*). Authenticity and accuracy (*Article 15*). Profiling restrictions (*Article 5, 52*).
Policy and compliance checks (Li et al., 2019; McIntosh et al., 2023a)	Automated policy drafting and checks (McIntosh et al., 2023a)	Introduction of non-compliant rules (Wang et al., 2023)	Automated information verification (*Article 60, 61, 64*). AI alignment with compliance (*Article 8, 9, 16, 24-27*).

[a] Including LLM-induced cybersecurity risks and inherent LLM risks.

- *Assessment of Controls for LLM Risks*: We listed LLM-specific controls using the rubric as a guideline. Then, we assessed the ability of each control to handle LLM risks.

(2) *Offensive Analysis:*

- *Framework Vulnerabilities to LLM Attacks*: We examined potential framework weaknesses using the rubric, focusing on areas LLMs could exploit. We also identified scenarios where LLMs may bypass regulations through generated content.
- *Holistic Framework Gap Analysis*: We comparatively evaluated overall LLM readiness using the rubric, spotlighting areas needing more LLM-specific provisions.

The mapping rubric, central to our methodology, has been developed to effectively pair specific framework attributes with established LLM opportunities and potential risks, as presented in Table 2. Due to the abstract principle-based nature of those frameworks, to appraise their LLM-readiness, we employed a qualitative binary "pass/fail" criterion, where "pass" indicates the framework is LLM-ready, and "fail" suggests the opposite. The use of a binary "pass/fail" criterion is justifiable on several grounds:

- *Nature of Qualitative Research*: Qualitative research is inherently interpretative, focusing on understanding the complexity and context of a subject rather than reducing it to numbers (Fujs et al., 2019). We believe a binary "pass/fail" system aligns with this interpretative nature, providing a clear, dichotomous outcome that reflects a framework's readiness without the false precision of a numerical score.
- *Complexity and Abstraction*: Cybersecurity frameworks are often characterized by their complexity and high level of abstraction, with provisions that cannot be easily quantified (Malatji et al., 2019). We believe a granular scoring system could oversimplify these provisions, potentially misrepresenting the nuanced assessment required for LLM integration. A binary system, by contrast, acknowledges this complexity and avoids the pitfalls of over-simplification.
- *Precedence in Qualitative Research*: Binary rating scales are not uncommon in qualitative evaluations, particularly in fields dealing with abstract concepts such as policy analysis and compliance assessments (Armenia et al., 2021; Pipyros et al., 2018). For instance, in the evaluation of regulatory frameworks, binary outcomes are often preferred to indicate adherence or non-adherence to standards, as they facilitate clear decision-making and action (Armenia et al., 2021; Malatji et al., 2019; Pipyros et al., 2018).
- *Clarity and Industry Standard*: The binary "pass/fail" assessment is widely recognized and utilized in the industry for compliance evaluations, providing a straightforward and decisive outcome (Leszczyna, 2021; Slapničar et al., 2022). This method is commonly paired with detailed feedback explaining the reasoning behind the outcome and offering advice for organizational improvement, ensuring transparency and actionable guidance for enhancing framework alignment.

Our approach, leveraging both human expertise and AI validation, provided a robust mechanism with a higher level of scrutiny and cross-validation than through human efforts alone, for ensuring the accuracy and integrity of our qualitative analysis. Our iterative process of human-LLM consensus served to balance the depth of human judgment with the breadth of the NLP analysis by LLM, as we integrated OpenAI's ChatGPT-4 and Anthropic's Claude AI into our analysis pipeline as follows:

- *Automated Validation Process*: Initially, the alignment of each framework with LLM capabilities and the EU AI Act provisions was independently assessed by the researchers. Then, to validate the analysis, the frameworks were processed through market-leading LLMs: ChatGPT-4, chosen for its advanced natural language prowess (McIntosh et al., 2023a), and Claude, selected for its market-recognized reliability and robustness to hallucinations (Toufiq et al., 2023). Both LLMs scrutinized the text of the frameworks and our initial assessments to pinpoint any discrepancies or aspects that may have been missed.

Table 3
Evaluation rubric for assessing cybersecurity frameworks' LLM readiness.

Criteria	Justification	Pass/Fail	Feedback
Governance structure	Frameworks must demonstrate robust and agile governance structures adaptable to AI challenges.		
Risk management capabilities	Ability to identify, assess, and mitigate risks introduced by LLMs.		
Compliance with relevant legislation	Alignment with existing and forthcoming AI regulations, including the EU AI Act.		
Ethical considerations	Inclusion of ethical guidelines for LLM deployment and management.		
Integration and augmentation potential	Frameworks should facilitate the beneficial integration of LLMs.		
Mitigation of LLM-specific risks	Provisions for addressing risks unique to LLMs, such as content hallucination.		
Transparency and human oversight	Requirements for clear LLM processes and human-centric decision-making.		

- *Consistency Checks*: The AI systems were tasked with verifying the consistency of applying our mapping rubric. They cross-referenced the identified LLM capabilities and risks with the provisions of the frameworks to ensure a thorough and unbiased application of the rubric criteria.
- *Discrepancy Resolution*: In instances where the AI findings diverged from the initial human assessment, the specific points of contention were re-evaluated. This step involved a detailed review of the relevant literature and framework documentation to resolve discrepancies, thereby refining the accuracy of our analysis.
- *Final Validation*: After achieving consistency between human and AI assessments, the final validation was conducted via expert review to confirm the robustness of the findings. This multi-stage process ensured a rigorous evaluation, minimizing subjective bias and enhancing the reliability of the qualitative analysis. It should be noted that this validation involved only two experts, which may limit the generalizability of the findings.

To provide a more justified and extensive rubric, we have formalized the evaluation process into a detailed table that breaks down the criteria for assessing the readiness of cybersecurity frameworks in relation to LLM integration. Our enhanced rubric maintains the binary "pass/fail" evaluation (common in cybersecurity compliance assessment) for clarity and compliance alignment, but with feedback for improvement, as detailed below in Table 3:

4. Study design and validation

This study adopted a thorough approach to assess cybersecurity frameworks, focusing on their engagement with LLMs within the advanced AI context, aiming to evaluate these frameworks' comprehensive preparedness for LLM-related threats and the forthcoming EU AI Act. The progression of our study is illustrated in Fig. 1. To ensure the maximum transparency, credibility, and reproducibility of our study, we exclusively used publicly available documents, specifically cybersecurity frameworks and the EU AI Act, without incorporating any customized data, to enable further scrutiny by other researchers for transparency and trustworthiness.

The steps of our study is as follows:

(1) *Cybersecurity Framework Selection:* Selecting the appropriate cybersecurity GRC frameworks for evaluation.
(2) *Framework Content Familiarization:* An extensive analysis of the cybersecurity framework content. The lead author, proficient in cybersecurity GRC, ensures that this foundational analysis is robust for subsequent steps.
(3) *Tri-Focal Analysis:*
   - *Defensive Analysis:* Evaluating the resilience of frameworks against LLM threats and their adaptability to harness LLM benefits, including:
     - Provisions of the framework against LLM-generated malicious content.
     - Mechanisms to harness the strengths of LLMs for enhanced cybersecurity.
   - *Offensive Analysis:* Unearthing potential vulnerabilities in the frameworks due to LLM interactions, including:
     - The ability of LLMs to produce misleading content.
     - Identifying gaps within the framework that LLMs might exploit.
   - *EU AI Act Readiness:* Evaluation of the cybersecurity framework in relation to the EU AI Act provisions.
(4) *Insight Synthesis:* Consolidation of the findings from the tri-focal analysis and synthesis of initial insights.
(5) *LLM Verification Method:* Construction of synthesized insights, combined with relevant sections from the selected cybersecurity framework and supporting evidence, as prompts for OpenAI's ChatGPT-4 and Anthropic's Claude, to independently:
   - Evaluate the validity of our initial analysis.
   - Highlight any potential gaps or oversights.

   Upon LLM feedback, if either ChatGPT-4 or Claude rejects our conclusions, or their reasoning does not align with our assessment, we restart our analysis. If they reject our analysis based on apparent hallucinations, we still restart, enhancing our evidence for better clarity. If both LLMs agree with our insights, we proceed to the expert review phase.
(6) *Expert Review:* Reflecting practices in specialized decision-making fields, our study leverages insights from a focused group of two independent subject matter experts in cybersecurity GRC, akin to the approach seen in LegalBench (Guha et al., 2022) and MultiMedQA (Singhal et al., 2023), where the complex nature of assessments of AI generated content in specialized decision-making fields often necessitates a smaller, highly specialized human evaluator pool to ensure depth and accuracy of analysis, when supported by well-defined rubrics and when the human feedback is provided for transparency and future scrutiny.
   - If the panel rejects, the lead author revisits content familiarization.
   - If accepted, progression to recommendation drafting ensues.

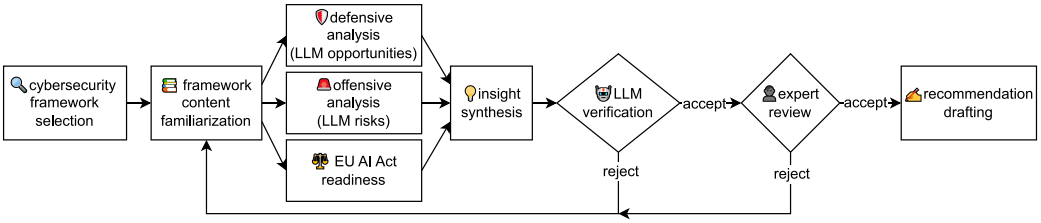


Fig. 1. Flowchart of the analysis process with LLM and GRC experts in the loop.

Table 4
Assessment of LLM opportunities, risks and EU AI Act compliance (✓: pass; ×: fail).

			NIST CSF 2.0	COBIT 2019	ISO 27001:2022	ISO 42001:2023
Process automation	LLM	Opportunities	×	×	✓	✓
		Risks	×	×	×	×
	EU AI Act readiness		×	✓	×	✓
Real-time analysis	LLM	Opportunities	✓	✓	✓	✓
		Risks	×	✓	×	×
	EU AI Act readiness		×	✓	×	✓
Data security and protection	LLM	Opportunities	✓	✓	✓	✓
		Risks	×	✓	×	✓
	EU AI Act readiness		✓	✓	✓	✓
Continuous monitoring and auditing	LLM	Opportunities	✓	✓	✓	✓
		Risks	×	×	×	✓
	EU AI Act readiness		✓	✓	✓	×
Incident response	LLM	Opportunities	✓	✓	✓	✓
		Risks	×	×	×	✓
	EU AI Act readiness		×	✓	×	×
Security awareness and training	LLM	Opportunities	×	✓	✓	✓
		Risks	×	×	✓	✓
	EU AI Act readiness		×	×	×	×
Policy and compliance checks	LLM	Opportunities	×	✓	✓	✓
		Risks	✓	×	✓	×
	EU AI Act readiness		×	✓	✓	×
Total marks	LLM	Opportunities	5/7	6/7	7/7	7/7
		Risks	2/7	2/7	2/7	4/7
	EU AI Act readiness		2/7	6/7	3/7	4/7

(7) *Recommendation Drafting:* Drafting actionable recommendations to strengthen the cybersecurity frameworks based on insights after the validation process.

5. Results

This section presents the findings from our analysis (Table 4), showcasing automation potential across frameworks, their capabilities for overseeing LLM-related risks, and identified gaps in readiness for LLM adoption and integration.

5.1. Validity of methodology and findings

The integrity of this study's methodology and the consequent findings hinge on a multi-layered validation process, integrating both computational and human expert assessments to ensure reliability and accuracy. Initially, the utilization of leading LLMs, ChatGPT-4 and Anthropic's Claude, provided a robust automated review, evaluating the consistency of framework assessments against LLM capabilities and emerging legal standards, such as the EU AI Act. This was complemented by an expert review, where two seasoned professionals in cybersecurity GRC provided valuable feedback, contributing to the study's credibility. Through this dual-faceted approach, we mitigated potential biases and enhanced the depth of our analysis, ensuring that our methodology not only aligns with current academic standards but also addresses the dynamic landscape of cybersecurity challenges posed by LLMs. The iterative validation process, combining AI insights with human expertise, underlines the robustness of our findings, contributing to the advancement of cybersecurity framework assessment methodologies in the age of LLM integration.

5.2. LLM opportunities

To systematically evaluate the potential of each framework for LLM automation and augmentation, we employed the mapping rubric to identify compatible processes and controls. Our assessment indicated all four frameworks accommodated some aspects of LLM capabilities, with CSF 2.0 offering the most comprehensive set of opportunities due to its breadth of technical and governance outcomes. However, high-level alignment did not preclude the need for additional LLM-specific provisions to ensure responsible and risk-aware integration.

NIST CSF 2.0 (rating 5/7). The NIST CSF 2.0 exhibits potential for integration with LLM technologies within the domains of "real-time analysis", "data security and protection", "continuous monitoring and auditing", and "incident response". These areas, though not explicitly addressing LLMs, provide a conducive framework for their application—risk assessment (*ID.RA*), data security (*PR.DS*), continuous monitoring (*DE.CM*), and incident management (*RS.MA*). However, it falls short in "process automation", "security awareness and training", and "policy and compliance checks", lacking specific references or provisions for LLM utilization, such as *Natural Language Processing* (NLP) for automation, adaptive security training modules, and automated policy drafting and compliance verification. These gaps suggest that while the framework may be adaptable to LLM integration, it currently does not offer explicit readiness in these critical areas.

COBIT 2019 (rating 6/7). COBIT 2019 presents notable alignment with LLM opportunities in the areas of real-time analysis (*APO12*), data security and protection (*APO01*, *BAI09*), continuous monitoring and auditing (*APO11*), incident response (*APO13*, *DSS04*, *DSS05*), security awareness and training (*BAI08*), and policy and compliance checks (*EDM01*, *EDM02*, *BAI01*, *BAI02*), though it has not specified the use of

LLMs within these provisions. The framework, however, does not pass in the domain of "process automation", as it lacks explicit guidance on integrating LLMs for NLP or automation. COBIT 2019 does not provide specific references to leveraging LLM capabilities for process automation, highlighting a gap that may need to be addressed to fully harness LLM opportunities in this aspect.

ISO 27001:2022 (rating 7/7). ISO27001:2022 demonstrates a readiness to leverage the capabilities of LLMs across various domains pertinent to cybersecurity and information security management. While the standard does not explicitly detail the integration of LLMs, its broad and thorough controls provide a robust foundation for the secure adoption and implementation of LLM technologies. Controls related to information security testing, system development, event logging, and incident management indicate an infrastructure that is conducive to incorporating LLMs into process automation, real-time analysis, and continuous monitoring. Additionally, the framework acknowledges the importance of security awareness and training, which can be enhanced through adaptive training modules potentially supported by LLMs. In "data security and protection", ISO27001:2022 requires the secure management of information throughout its lifecycle, which is essential for LLMs handling sensitive data. Continuous monitoring and auditing controls imply a supportive environment for LLMs' continuous learning processes, ensuring that their evolving capabilities remain within the realm of secure operations. Incident response controls align well with the use of LLMs for automated detection and reporting, facilitating timely and effective incident management. Overall, while ISO27001:2022 is not LLM-specific, its flexible and technology-agnostic approach allows it to remain relevant as new technologies emerge, indicating a strong potential for integration with LLM opportunities. The standard's focus on information security management aligns with the inherent needs of LLMs, especially regarding the protection of data, which they process and generate. The framework provides extensive coverage that can be interpreted to support the secure introduction and utilization of LLMs within the constraints of its control sets.

ISO 42001:2023 (rating 7/7). ISO 42001:2023, while not explicitly designed for LLM integration, offers a framework that can support LLM capabilities in several key areas. For "process automation", its focus on continuous improvement and risk management aligns with the needs for natural language understanding and generation, potentially offering a supportive environment. In "real-time analysis", the standard's emphasis on continual improvement and adaptability may facilitate real-time data processing. For "data security and protection", ISO 42001's comprehensive risk management approach could be conducive to integrating data analytics, image recognition, and generation. The standard's focus on "continuous monitoring and auditing" aligns well with the needs for continuous learning in LLMs. In "incident response", the structured approach to risk assessment and treatment may support automated incident detection and reporting. The "security awareness training" aspect could benefit from the standard's emphasis on awareness and competence, potentially enabling the integration of adaptive training modules. Lastly, in "policy checks and compliance", ISO 42001's structured approach to managing AI systems and risks may align with the requirements for automated policy drafting and checks, though explicit provisions for LLMs are not detailed. Overall, while ISO 42001:2023 does not specifically address LLMs, its principles and focus on AIMS suggest a supportive framework for their integration, warranting further exploration and application to determine its full compatibility.

5.3. LLM risks

To systematically evaluate the provisions of each framework for governing LLM-associated risks, we employed the mapping rubric to identify relevant controls and requirements.

NIST CSF 2.0 (rating 2/7). NIST CSF 2.0 offers firm guidance within "Policy and compliance checks" via the *GV.PO* category, requiring robust policies for safeguarding data and technology. This, however, marks the extent of its explicit coverage of LLM-related risks, particularly in the specialized areas of content generation, real-time bias correction, data protection specific to LLM technology, continuous LLM data oversight, targeted LLM incident response, and LLM-focused security education. The framework does not satisfactorily address the intricacies of LLM risk in "Process Automation", failing to offer specific strategies for the perils arising from LLM-generated content. The "Real-time Analysis" component, while presenting relevant categories, falls short in providing concrete measures for the unique temporal and bias-related challenges associated with LLM outputs. Within "Data Security and Protection", NIST CSF 2.0 establishes a broad defense but stops short of probing into the advanced threats LLMs pose, such as data breaches and deceptive visual content. Its lack of detailed guidance on the relevance and security of training data indicates a gap in "Continuous Monitoring and Auditing", essential for LLM-specific oversight. "Incident Response" protocols are not adequately calibrated for the specific issues of LLM misclassifications, potentially leading to ineffective response actions. Although PR.AT-02 requires specialized role training, those implementing PR.AT of NIST CSF 2.0 should explicitly ensure training covers LLM-specific knowledge, as LLMs introduce unique issues and challenges not present in existing cybersecurity knowledge.

COBIT 2019 (rating 2/7). COBIT 2019 meets LLM risk readiness criteria in the areas of "Real-time analysis" and "Data security and protection", with relevant provisions being *EDM03* and *DSS05* respectively. These sections provide a foundation for addressing risks associated with real-time processing and data protection, which could extend to encompass the unique challenges posed by LLMs. Conversely, COBIT 2019 fails to adequately address LLM risk readiness in the domains of "Process automation", "Continuous monitoring and auditing", "Incident response", "Security awareness and training", and "Policy and compliance checks". Its general IT governance and management objectives do not sufficiently cover the specific risks associated with automated content generation by LLMs, nor do they ensure that risk responses are specifically tailored for the challenges posed by LLMs, such as incident misclassification or the introduction of non-compliant rules. Furthermore, there is an absence of detailed guidance for the management and monitoring of LLM training data and the subtle needs of AI/ML-specific employee training content, which is crucial for maintaining an informed and prepared workforce in the face of evolving LLM risks.

ISO27001:2022 (rating 2/7). ISO27001:2022 has demonstrated a foundational readiness for "Security awareness and training" and "Policy and compliance checks", under provisions *A.7.2* for fostering security knowledge, and *A.18.1* and *A.18.2* for policy management, yet these lack explicit directives for LLM-specific risks. The framework has not adequately covered "Process automation", with no tailored controls for automation risks inherent to LLMs in its Annex A, notably absent in sections addressing separation of environments (*A.12.4.1*). For "Real-time analysis", it has fallen short, missing explicit consideration for LLM-induced biases and latency. The "Data security and protection" provision, although robust in its scope (*A.8.2.3* and *A.14.1.2* among others), has failed to specifically safeguard against LLM-related risks like visual deception and forensic unreliability. Its "continuous monitoring and auditing" aspect has lacked directives on ensuring the ongoing relevance and integrity of the training data. In "Incident response", its general incident management controls (*A.16.1*) have not directly addressed the unique challenge of LLM misclassification risks. Consequently, while ISO27001:2022 has established a broad security and compliance framework, it still requires significant enhancement to directly confront the unique challenges posed by LLM technologies.

ISO 42001:2023 (rating 4/7). ISO 42001:2023 has demonstrated a mixed readiness for managing LLM-related risks. In our assessment, the standard passed in the domains of "Data security and protection", "Continuous monitoring and auditing", "Incident response", and "Security awareness training". It provided comprehensive guidelines for data management (including privacy and security implications), AI system logging, and promoting security knowledge (Controls B.7.2, B.7.3, and A.7.2). However, it failed in the areas of "Process automation", "Real-time analysis", and "Policy and compliance checks". While it addressed general AI system risks, specific references to managing misleading content generation, real-time biases, or the introduction of non-compliant rules in AI systems (including LLMs) were lacking. The closest relevant provisions included data quality requirements and ensuring the responsible use of AI systems (Controls B.7.4, B.9.2), yet these did not fully cover the complex risks posed by LLMs, such as automated decision-making biases or the unique challenges of real-time AI system responses. This gap indicated a need for more detailed and explicit risk management strategies for LLM technologies within the standard.

5.4. EU AI Act readiness

The EU AI Act introduces specific provisions for organizations implementing LLMs, which are distinct from those for regular AI systems due to the unique capabilities and risks associated with LLMs:

- **Risk Management Systems (*Article 9*)**: LLMs can process and generate content at scale, increasing the risk of widespread misinformation or data manipulation.
  Our recommendation: Implement systems to assess, document, and minimize such cybersecurity risks.
- **Mandatory Cybersecurity Testing (*Article 15*)**: The complexity and depth of LLMs may harbor hidden vulnerabilities.
  Our recommendation: Require extensive testing for vulnerabilities and data integrity before deploying LLMs.
- **Transparency Obligations (*Article 13*)**: LLMs' "black box" nature makes understanding their decision-making processes challenging.
  Our recommendation: Mandate documentation on high-risk LLMs' capabilities, limitations, and security measures for clarity.
- **Post-Market Monitoring (*Article 61*)**: The evolving nature of LLMs means new risks can emerge after deployment.
  Our recommendation: Require continuous monitoring for cybersecurity issues.
- **Record-Keeping (*Article 11-12*)**: The adaptive learning of LLMs necessitates detailed records of their design, risk assessments, and evaluations.
  Our recommendation: Make such records accessible for authority review to ensure ongoing compliance.
- **Reporting Obligations (*Article 62*)**: Given LLMs' potential impact, significant cyber incidents must be reported to authorities.
  Our recommendation: Ensure accountability and rapid response to threats posed by LLMs, and prompt reporting of serious incidents and malfunctioning to regulatory bodies.
- **Appointment of Cybersecurity Officers (*Article 17*)**: LLMs require specialized oversight due to their complex nature.
  Our recommendation: Appoint qualified cybersecurity officers to oversee LLM security compliance.
- **Fines for Non-Compliance (*Article 71*)**: Non-compliance with LLM-specific cybersecurity requirements can result in financial penalties.
  Our recommendation: Adhere to the heightened security needs of LLMs.

Given the frameworks analyzed in relation to the EU AI Act provisions, none explicitly include AI-specific stipulations. However, they exhibit varying degrees of implicit alignment with the Act requirements, with some fulfilling numerous provisions without necessitating major amendments.

NIST CSF 2.0 (rating 2/7). The NIST CSF 2.0 has demonstrated alignment with the EU AI Act in areas of "data security and protection" (*PR.DS*) and "continuous monitoring and auditing" (*DE.CM*), emphasizing robust protection against data breaches, ensuring AI data quality and integrity, and fostering continuous monitoring with periodic AI assessment. The "process automation" provision of NIST CSF 2.0 is not EU AI Act ready, because it lacks specific requirements related to transparency, oversight, and compliance of automated processes, and may not fully address the AI-specific requirements. Its "real-time analysis" provision does not adequately cater to real-time safeguards and unbiased AI system responsiveness. Its "incident response" provision fails to specify rapid AI-driven incident recognition and strategies for AI misclassification risk mitigation. Its "security awareness and training" provision is deficient in terms of AI transparency in training, authenticity and accuracy of AI-focused training information, and profiling of employees. Lastly, the "policy and compliance checks" provision is not comprehensive in addressing automated information verification and direct guidelines for AI alignment with compliance.

COBIT 2019 (rating 6/7). COBIT 2019 has exhibited strong EU AI Act readiness across several provisions, with its governance and management objectives addressing requirements for process automation, real-time risk management, data security and protection, continuous monitoring and auditing, incident response, and policy and compliance checks, emphasizing transparency, oversight, real-time safeguards, robust data protection, continuous evaluation, agile incident responses, and compliance alignment. While its "security awareness and training" provision promotes comprehensive training and awareness related to ethics, transparency, and appropriate use of information, it may lack in-depth AI-specific considerations in alignment with the EU AI Act, potentially missing direct provisions on employee profiling in an AI context and specific guidelines on the authenticity and accuracy of AI-focused training information. It is acknowledged that expecting direct AI-specific references from a standard not primarily focused on AI might be unrealistic, highlighting a potential area for future updates to address emerging AI risks comprehensively.

ISO 27001:2022 (rating 3/7). ISO 27001:2022 has demonstrated a degree of EU AI Act readiness in its provisions related to data security and protection, continuous monitoring and auditing, and policy and compliance checks, emphasizing robust guidelines against data breaches, continuous evaluation of information security controls, and comprehensive policy and compliance assessment. However, there are areas of concern: its "process automation" provision is not EU AI Act ready, as it lacks specific guidance around transparency, oversight, and compliance for automated processes in the AI context. Its "real-time analysis" provision does not fully address the requirements around unbiased AI system responsiveness and real-time safeguards. Its "incident response" provision, while robust in general incident management, does not target AI-driven recognition or misclassification risks, necessitating further guidelines to handle challenges posed by AI systems. Its "security awareness and training" provision lacks direct provisions for AI-specific issues such as AI transparency in training and employee profiling restrictions.

ISO 42001:2023 (rating 4/7). ISO 42001:2023 has demonstrated considerable readiness in several aspects of EU AI Act compliance, but with areas needing further enhancement. In "Process automation", it aligns well with transparency and human oversight requirements (Article 52–53, 61, 63) through its focus on risk treatment and effectiveness verification (*Clauses 6.1.3 and 6.1.4*). For "Real-time analysis", ISO 42001:2023 partially meets the criteria of real-time safeguards and unbiased AI responsiveness (Article 5, 9, 62, 65) through its provisions for monitoring and measuring AIMS performance (*Clause 9.1*). The framework effectively addresses "Data security and protection" with robust protection against data breaches and integrity of AI data (Article 15, 70, Annex IV-2, VII-4.3) by ensuring effective internal audits (*Clauses 9.2*

and 9.2.1) and top management reviews (*Clause 9.3*). However, gaps are observed in "Continuous monitoring and auditing" and "Incident response", lacking direct provisions for periodic AI assessments and AI-driven incident recognition, despite general clauses on corrective actions and nonconformity management (*Clause 10.1*). "Security awareness training" is partially covered, addressing AI transparency (*Clause 7.4*), but lacking specifics on authenticity, accuracy, and profiling restrictions. In "Policy and compliance checks", ISO 42001:2023 excels in automated information verification and AI compliance (Article 60, 61, 64, 8, 9, 16, 24–27), thanks to its comprehensive framework for AIMS implementation, maintenance, and continuous improvement, providing a structured approach to AI governance and compliance.

5.5. Gap analysis

Our assessment has revealed key insights into the readiness of the frameworks, including the latest ISO 42001:2023, for LLM integration along two dimensions: automation potential and risk oversight. A comparative analysis highlights crucial gaps that need to be addressed as summarized in Table 5. While the NIST CSF 2.0 offers extensive automation potential, it lacks explicit LLM risk oversight. COBIT 2019 facilitates high-level automation opportunities but requires more granular technical controls for LLM-specific risks. ISO 27001:2022 provides a solid foundation for human-centered LLM adoption, yet needs augmentation for full automation potential. ISO 42001:2023, although not specifically targeting LLMs, shows promise in several domains such as process automation and data security, but requires further refinement in areas like real-time analysis and policy compliance for LLM applications. Our findings emphasized the necessity for a multi-dimensional approach as cybersecurity frameworks evolve to support LLM integration. This involves addressing both automation opportunities and strengthening risk oversight specific to LLM technologies. All frameworks, including ISO 42001:2023, while showing automation readiness, need enhancement to implement LLMs securely. This could be through oversight processes for NIST CSF 2.0, technical validations for COBIT 2019, automation-focused provisions for ISO 27001:2022, and more explicit LLM-related guidelines in ISO 42001:2023. The findings reiterate the need for frameworks to adopt a multi-dimensional view (Fig. 2), considering automation potential, oversight, and EU AI Act readiness, to support the integration of LLMs into cyber risk management programs. Consequently, we recommend caution if organizations wish to adopt any existing cybersecurity GRC framework without developing a false sense of security in adopting LLM opportunity readiness, LLM risk readiness, and EU AI Act readiness.

5.6. Common weakness in addressing LLM hallucination

The oversight of "misleading content generation", colloquially known as LLM "hallucination" (Ji et al., 2023), emerged as a shared deficiency across the four frameworks. Understanding this oversight requires a deep dive into the complex nature of hallucination and its implications for cybersecurity. Existing literature defines LLM hallucination as text generated by LLMs that contains factual inconsistencies, contradictions, or content that diverges from human cultural norms and expectations, even when being coherent and seemingly realistic (Ji et al., 2023; McIntosh et al., 2023b, 2024a). This definition, while capturing the essence of the problem, falls short in addressing the subjective nature of hallucinations. To address this, our definition of LLM hallucination has been expanded to include not only factually inconsistent outputs, but also those outputs that might be culturally incoherent or diverge from mainstream human values and expectations. This broader perspective recognizes that identifying a hallucination is not merely a task of matching facts but also involves the application of cultural and value-based perspectives, emphasizing the importance of human-centric assessments. From a cybersecurity perspective, we believe the implications of LLM hallucinations are multi-dimensional:

(1) *Misinformation and Disinformation*: LLM hallucination poses risks of propagating misinformation and disinformation, both general and culturally specific. In cybersecurity processes, relying on culturally inappropriate or factually incorrect information can lead to flawed decision-making.
(2) *Integrity of Data*: Compromised data quality and reliability due to LLM hallucinations can lead to erroneous conclusions in cybersecurity decision-making.
(3) *Diverse Stakeholder Impact*: Culturally and ideologically incoherent hallucinations could lead to misinterpretations or be considered offensive, affecting collaboration and trust among stakeholders of different backgrounds.
(4) *Decision-making Complexity*: Complex risk assessments in cybersecurity are further complicated by hallucinated or incorrect LLM outputs, which could lead to decision-making paralysis.
(5) *Human-AI Dynamics*: The rise of LLM-driven tools in cybersecurity necessitates harmony between human decisions and LLM recommendations, which is negatively affected by introducing LLM outputs that humans find difficult to interpret or trust.

The absence of adequate provisions to address LLM hallucinations in those frameworks has highlighted a broader issue: while these standards are forward-looking, they might not yet be equipped to address the complex challenges posed by emerging technologies like LLMs, and require frequent and robust updates to integrate specific checks and controls to identify, manage, and mitigate the risks posed by LLMs.

6. Discussion

The integration of LLMs into the realm of cybersecurity frameworks presents a multi-dimensional challenge that intersects both technological capabilities and governance oversight. In this section, we present some of our insights for discussion.

6.1. The necessity of multi-pronged framework evolution

Our findings have underlined the need for cybersecurity frameworks to undergo a multifaceted evolution. Just as previous research identified gaps in handling emerging technologies like cloud computing (Darraj et al., 2019) and general AI (Kaur et al., 2023), our analysis extends these insights to include LLMs. Current frameworks have shown foundational readiness, but require enhancements for fully harnessing LLM capabilities and effectively managing their inherent risks. This evolution is in line with the dual-track approach that advocates for balancing innovation and oversight in emerging technology adoption (Asad et al., 2023; Ukil et al., 2023). ISO 42001:2023, while designed for and comprehensive in AI system management, still demonstrated areas needing refinement, particularly in addressing LLM-specific risks and compliance checks. This echoed the necessity of framework evolution to encompass adaptive controls tailored to the unique opportunities and challenges posed by LLMs (Ji et al., 2023; Darraj et al., 2019). Similarly, NIST CSF 2.0 and ISO 27001:2022, despite their foundational strengths, require updates to capitalize fully on LLM integration and risk management. This includes NIST CSF 2.0 enhancing its risk oversight, particularly for mitigating misleading content, and ISO 27001:2022 expanding its guidance for leveraging automation in training and monitoring. COBIT 2019's need for additional technical provisions for LLM-specific risk management reiterates this trend.

Cybersecurity GRC frameworks must refine their guidelines on training, system development, automation, and policy compliance, now also considering the guidelines established by ISO 42001:2023, to fully unlock LLMs' potential in transforming cybersecurity operations (Akande et al., 2023; Abie, 2019; Markopoulou et al., 2019). The integration of human-centered collaboration, involving a diverse range of stakeholders, into framework design and implementation is

Table 5
Comparative gap analysis of cybersecurity frameworks in LLM readiness.

Framework	LLM opportunities	LLM risks	EU AI Act
NIST CSF 2.0	Incomplete provisions for process automation, security training, and policy compliance specific to LLM; lacks references to LLM for natural language processing, adaptive security training, and automated policy compliance.	Insufficient measures for LLM-generated content risks, real-time bias, advanced data breach threats, LLM data oversight, and tailored LLM incident response.	Partially aligns with data security and monitoring but lacks readiness in process transparency, oversight, unbiased responsiveness, AI-driven incident response, and AI-specific training and compliance.
COBIT 2019	Omits explicit guidance on LLM integration for process automation; though covering various domains, it misses out on natural language processing and automation specific to LLM.	Does not cover LLM automated content generation risks; lacks LLM-tailored risk responses and specific AI/ML training for workforce.	Exhibits substantial readiness; however, its security training lacks depth in AI-specific considerations, employee profiling in AI, and training authenticity as per EU AI Act.
ISO 27001:2022	While broad, does not detail LLM integration; controls need interpretation to support LLM application in process automation, real-time analysis, and continuous monitoring.	Provides foundation for security training and policy compliance but lacks explicit LLM risk directives; inadequate for LLM automation, real-time analysis, and incident misclassification risks.	Addresses some EU AI Act requirements but is not ready in providing guidance for transparency in AI process automation, real-time unbiased AI responses, and AI-driven incident management specifics.
ISO 42001:2023	Supports LLM capabilities in process automation, real-time analysis, data security and protection, and more, through its focus on AIMS; however, does not explicitly detail LLM integration.	Effective in certain domains like data security and incident response, but lacks explicit strategies for LLM-specific risks such as misleading content generation and real-time biases.	Shows considerable readiness in some aspects of the EU AI Act, aligning well with transparency and oversight requirements, but has gaps in continuous monitoring and AI-driven incident response specifics.

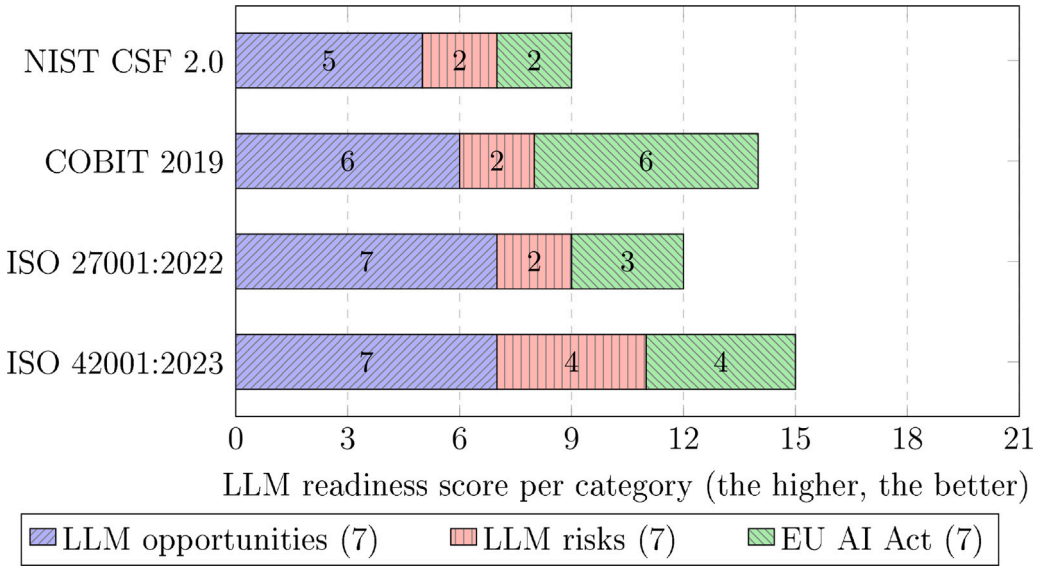


Fig. 2. Comparison of cybersecurity GRC frameworks in LLM readiness.

essential. This ensures that frameworks reflect the priorities of both technology leaders and ethical oversight experts, thereby enhancing real-world efficacy (Paskauskas, 2022; Malaivongs et al., 2022; Manuel et al., 2022). Our findings, along with recent scholarship, advocate for a re-examination of risk paradigms in light of AI advancements, acknowledging that current models may not fully account for the dynamic nature of technologies like LLMs (Katina and Keskin, 2021; Manuel et al., 2022). Thus, organizations must adopt a proactive stance in framework implementation, anticipating and aligning with the evolving capabilities and risks of LLMs to maintain cybersecurity effectiveness (Karie et al., 2021; Kaur et al., 2023; Gupta et al., 2023; Ukil et al., 2023).

6.2. The significance of continuous evolution and version control

Our analysis found that frameworks must undergo rapid yet robust evolution to address emerging technologies. However, version control is crucial to ensure organizational adoption keeps pace with framework revisions. The identified gaps in the latest versions of the NIST CSF, COBIT, and ISO frameworks concerning LLM oversight underscore concerns about the frameworks' agility in keeping up with AI advances, highlighted in studies on framework modernization challenges (Gourisetti et al., 2020; Hitchcox, 2020). While the pace of technological change is a reasonable challenge, it necessitates urgent version updates coupled with effective transition planning. This is to

minimize prolonged lapses in readiness, a critical point underlined by researchers studying the adaptability and responsiveness of cybersecurity frameworks to emerging threats (Gourisetti et al., 2020; Hitchcox, 2020). Our findings complement this discourse by demonstrating that the limitations of existing frameworks extend beyond operational aspects; they are conceptual, often failing to incorporate anticipatory governance necessary for technologies like LLMs (Katina and Keskin, 2021; Manuel et al., 2022). This concept echoes the necessity for organizations not only to update their GRC frameworks more frequently, but also to integrate forward-looking approaches that can keep pace with AI innovation (Darraj et al., 2019; Kaur et al., 2023). Therefore, we advocate for strategic version control to ensure that updated frameworks permeate through organizational infrastructure in a timely manner.

Our findings have thus highlighted that continuous evolution of frameworks must be complemented by responsible version release and adoption within organizations. All four frameworks need enhanced evolution to address technological changes rapidly, paired with strategic organizational implementation of updated versions. This emphasizes the significance of agile development and timely adoption for cybersecurity frameworks to remain relevant against emerging technologies. Balancing evolution and adoption is key for frameworks to continue fulfilling their vital role as cyber risk navigation tools in a climate of unrelenting change (Angelini et al., 2017).

6.3. Strengthening provisions for LLM hallucination risks

Our investigation has exposed a lack of human oversight in the management of LLM hallucination risks within those 4 frameworks investigated, a concern that may be prevalent across other cybersecurity frameworks. This absence of controls aligns with the discourse in prior works advocating for risk management strategies tailored to AI's unique threats (Darraj et al., 2019; Karie et al., 2021; Kaur et al., 2023), which our focus on LLM hallucination risks specifically seeks to advance. The deceptive nature of LLM hallucinations, which can be both subtle and overt, exacerbates these risks, especially when paired with human complacency or insufficient human oversight (Ji et al., 2023). Those frameworks have been found to lack LLM-specific controls, particularly against the propagation of misleading content, which poses risks such as misinformation spread, data integrity breaches, stakeholder misalignment, decision-making disruption, and compromised human-AI collaboration. Prior research has underscored the necessity for cybersecurity measures that specifically address the unique threats posed by AI, advocating for a shift in risk management strategies to further encompass LLM's distinct threat profile (Darraj et al., 2019; Karie et al., 2021; Kaur et al., 2023; Argyridou et al., 2023; Bozkus Kahyaoglu and Caliyurt, 2018). To bridge these gaps, organizations must adopt a strategic approach that recognizes LLM as an autonomous entity within the threat landscape, and tactically integrate LLM risk scenarios into their cybersecurity exercises to fine-tune their response to LLM-specific threats (Katina and Keskin, 2021; Manuel et al., 2022; Syafrizal et al., 2020).

To enhance the human oversight of LLM hallucination risks, several measures are recommended for incorporation into those frameworks. Instituting mandatory hallucination identification processes, such as confidence scoring and uncertainty quantification, can preemptively detect misleading LLM outputs (Huang et al., 2023; Schuster et al., 2022). Implementing human validation checkpoints ensures critical human oversight in the review of LLM outputs, while mandated transparency around LLM training data and model functionality aids in discerning unreliable outputs (Gupta et al., 2023; McIntosh et al., 2023a). Continuous bias testing is also essential, uncovering and correcting distortions in LLM knowledge bases that may lead to hallucinations (Meskó and Topol, 2023). By embedding these targeted provisions, the frameworks can significantly bolster organizational defenses against LLM hallucinations, offering a comprehensive model for security standards that aim to integrate LLMs and manage their complex vulnerabilities effectively.

6.4. Limitations of the study

This study has provided insights into how key cybersecurity frameworks integrate advanced AI systems like LLMs, yet it has limitations. It acknowledges that, apart from ISO 42001:2023, the selected frameworks do not specifically address AI adaptation and governance, which could limit their applicability to emerging AI technologies. The qualitative content analysis's inherent subjectivity risks bias, despite efforts to mitigate this through validation with LLMs with NLP and human GRC experts. Such validation cannot ensure statistically generalizable conclusions and limits the findings to the examined contexts and datasets. Our focus on four principal frameworks, while thorough, excludes other standards that could offer additional comparative insights. The reliance on publicly available documents ensures transparency, but may miss the subtleties of dynamic stakeholder interactions. The study's scope, focusing on the readiness of specific cybersecurity frameworks without incorporating a comprehensive legal analysis of the EU AI Act's limitations, presents a constraint, reflecting a broader challenge in anticipating the impact of future legislative amendments on AI and cybersecurity practices. Despite these constraints, the study lays groundwork for future research into LLMs' influence on cybersecurity governance, risk, and compliance, underscoring the challenges and opportunities of these AI systems. The use of a "pass/fail" approach, while common in compliance assessments, oversimplifies the compliance continuum. The abstract and principle-based nature of cybersecurity GRC frameworks adds to the analysis's subjectivity. Nevertheless, including both LLMs and human experts aims to minimize interpretive discrepancies.

6.5. Future directions and implications

Future work could utilize mixed methods, expanded scope, and cross-disciplinary perspectives to further enrich understanding of how leading cybersecurity frameworks can continue evolving to support the safe, ethical and effective adoption of rapidly emerging technologies like AI. Findings would inform development of agile, holistic and evidence-based standards and programs for cybersecurity governance, risk management and compliance, ensuring frameworks remain relevant and effective in the face of new AI challenges. Here are some suggestions for future work based on the limitations and findings of this study:

- Launch a survey targeting cybersecurity professionals to statistically quantify the readiness and identify gaps in LLM integration across sectors, aiming for data that can validate findings and guide framework updates.
- Include a broader range of emerging and specialized cybersecurity frameworks like NIST SP 800-53 in the scope to achieve a more detailed comparative analysis and understand sector-specific requirements for AI systems.
- Undertake case studies to observe the practical application of frameworks in organizations, focusing on effectiveness and real-world challenges in managing AI risks, with an emphasis on the use of LLMs.
- Extend the investigation to other cutting-edge technologies, including IoT, for a holistic view of how current frameworks can adapt to the broader technological landscape.
- Evaluate the responsiveness of frameworks to the rapidly evolving landscape of AI-enhanced threats, emphasizing the need for swift integration of new protective measures.
- Conduct a thorough examination of how cybersecurity frameworks currently align with not just the EU AI Act, but also other emerging legislations and standards in AI ethics and security.
- Facilitate focus groups or utilize Delphi methods to dynamically extract expert insights, allowing for a richer, contextually nuanced understanding of framework application in the era of generative AI.

- Probe the potential benefits of integrating cybersecurity frameworks with other disciplinary perspectives, such as ethics and psychology, to create a more robust approach to AI challenges.

7. Conclusion

This study has conducted a comprehensive evaluation of four leading cybersecurity frameworks – NIST CSF 2.0, COBIT 2019, ISO 27001:2022, and the recently introduced ISO 42001:2023 – in the context of their readiness for integrating and governing the rapidly evolving domain of LLMs. Employing a detailed qualitative approach that includes content analysis, AI validation, and expert reviews, we unearthed a varying degree of alignment of these frameworks with the capabilities and risks associated with LLMs, indicating both promising potentials for integration and significant gaps in risk management. ISO 27001:2022 demonstrated strengths in human-centric validation for LLM outputs, reflective of its comprehensive approach to information security management. However, it became evident that all frameworks, including the NIST CSF 2.0 and COBIT 2019, necessitate further refinement to fully embrace the opportunities and address the cybersecurity risks introduced by LLMs from both technical and governance perspectives. Interestingly, ISO 42001:2023, despite being the most recent framework specifically designed for AI management, was found to have certain limitations in fully addressing the unique challenges and opportunities presented by LLM commercialization. While ISO 42001:2023 emerged as a frontrunner in our comparative analysis, its principles and guidelines, primarily tailored for general AI management, did not fully encapsulate the specific subtleties and requirements of LLM technologies. This gap highlighted the necessity for even the most contemporary frameworks to undergo continuous evolution, ensuring that they are not only in tune with general AI advancements but also adequately responsive to the distinct characteristics of LLMs, particularly the phenomenon of 'hallucination' or misleading content generation. This study emphasizes the urgent need for the modernization of mainstream cybersecurity standards to effectively govern these emerging threats. A critical takeaway is the requirement for cybersecurity frameworks to incorporate LLM-tailored controls that emphasize transparency, human validation, bias testing, and continuous monitoring. As the landscape of cybersecurity and AI continues to evolve, our findings call for a proactive and dynamic approach in the development and adaptation of cybersecurity frameworks, ensuring their relevance and efficacy in the age of advanced LLMs and beyond.

Abbreviations

AI	Artificial Intelligence
COBIT	Control Objectives for Information and Related Technologies
CSF	Cybersecurity Framework
EU	European Union
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformers
GRC	Governance, Risk and Compliance
ISO	International Organization for Standardization
LLM	Large Language Model
ML	Machine Learning
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing

CRediT authorship contribution statement

Timothy R. McIntosh: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Teo Susnjak:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis. **Tong Liu:** Writing – review & editing, Validation, Methodology, Conceptualization. **Paul Watters:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Dan Xu:** Writing – review & editing. **Dongwei Liu:** Writing – review & editing. **Raza Nowrozy:** Writing – review & editing. **Malka N. Halgamuge:** Writing – review & editing, Writing – original draft, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix. A proposed year-long plan for transition

Based on our insights, we also propose a year-long plan for responsible organizations in charge of those cybersecurity GRC frameworks (*i.e.*, NIST CSF 2.0, COBIT 2019, ISO27001:2022, and ISO42001:2023), divided into four quarters, with the objective of updating those four cybersecurity GRC frameworks for the integration and regulation of LLMs. An illustrated Gantt Chart is provided in Fig. A.3. While acknowledging the complexity of revising cybersecurity GRC frameworks, and the possibility that such a task may extend beyond one year, the impending passage of the EU AI Act within the forthcoming year (2024) necessitates an expedited timeline. Consequently, we have designed this roadmap as a one-year project to address the criticality of the situation. Nevertheless, the project committee may exercise discretion to scale the pace of updates as required.

- Quarter 1
  - Establish an interdisciplinary task force with expertise in cybersecurity, AI, and legal compliance.
  - Conduct a comprehensive gap analysis to determine current framework deficiencies with respect to LLM integration and EU AI Act requirements.
  - Develop a revision strategy for each framework, focusing on automation opportunities, risk governance, and regulatory compliance.
- Quarter 2
  - Begin framework revision with a focus on identifying and mitigating LLM hallucination risks.
  - Institute processes for enhanced transparency, validation mechanisms, and bias testing specific to LLM usage.
  - Initiate consultations with industry and academia to ensure the practicality and relevance of the proposed revisions.
- Quarter 3
  - Implement version control protocols to manage the transition to updated framework versions efficiently.
  - Complete and pilot revised draft frameworks within selected organizations for real-world testing and feedback.
  - Revise training programs and certification requirements to include LLM-specific content.
- Quarter 4
  - Finalize framework revisions, incorporating feedback from the pilot phase and ensuring alignment with the EU AI Act.
  - Release the updated frameworks with comprehensive transition guidelines for organizations.
  - Launch a global awareness campaign to inform stakeholders of the new revisions and encourage widespread adoption.

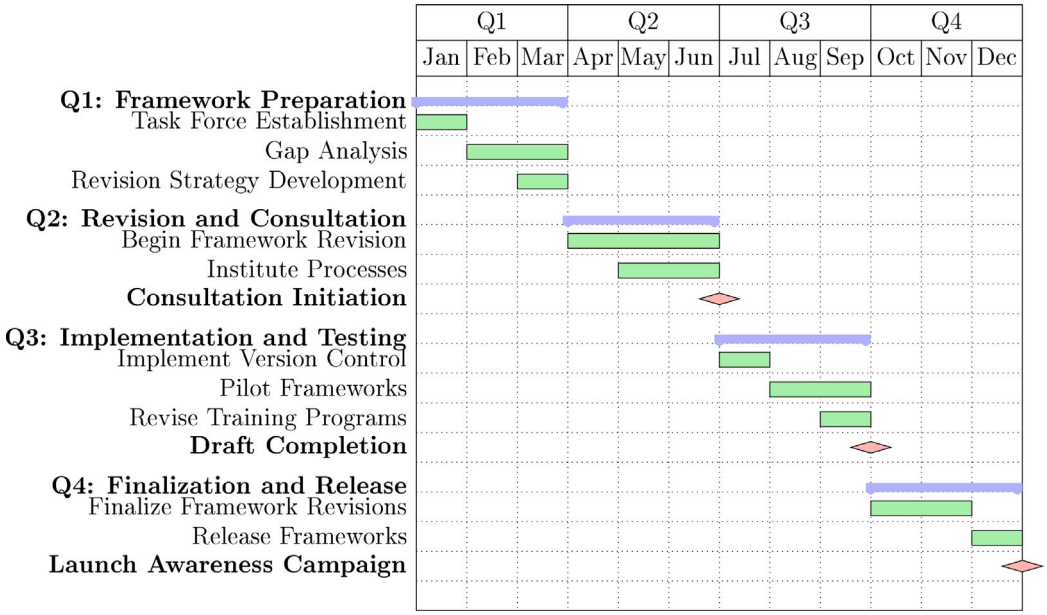


Fig. A.3. 1-Year plan Gantt chart to revise cybersecurity GRC frameworks.

References

Abie, H., 2019. Cognitive cybersecurity for CPS-IoT enabled healthcare ecosystems. In: 2019 13th International Symposium on Medical Information and Communication Technology. ISMICT, IEEE, pp. 1–6.

Akande, A.J., Foo, E., Hou, Z., Li, Q., 2023. Cybersecurity for satellite smart critical infrastructure. In: Emerging Smart Technologies for Critical Infrastructure. Springer, pp. 1–22.

Alromaih, A., Ismail, Y., Elmedany, W., 2022. Continuous compliance to ensure strong cybersecurity posture within digital transformation in smart cities. In: 6th Smart Cities Symposium. SCS 2022, Vol. 2022, IET, pp. 464–479.

Angelini, M., Lenti, S., Santucci, G., 2017. Crumbs: a cyber security framework browser. In: 2017 IEEE Symposium on Visualization for Cyber Security. VizSec, IEEE, pp. 1–8.

Argyridou, E., Nifakos, S., Laoudias, C., Panda, S., Panaousis, E., Chandramouli, K., Navarro-Llobet, D., Mora Zamorano, J., Papachristou, P., Bonacina, S., 2023. Cyber hygiene methodology for raising cybersecurity and data privacy awareness in health care organizations: Concept study. J. Med. Internet Res. 25, e41294.

Armenia, S., Angelini, M., Nonino, F., Palombi, G., Schlitzer, M.F., 2021. A dynamic simulation approach to support the evaluation of cyber risks and security investments in SMEs. Decis. Support Syst. 147, 113580.

Asad, U., Khan, M., Khalid, A., Lughmani, W.A., 2023. Human-centric digital twins in industry: A comprehensive review of enabling technologies and implementation strategies. Sensors 23 (8), 3938.

Atrinawati, L., Ramadhani, E., Fiqar, T., Wiranti, Y., Abdullah, A., Saputra, H., Tandirau, D., 2021. Assessment of process capability level in university XYZ based on COBIT 2019. In: Journal of Physics: Conference Series. Vol. 1803, IOP Publishing, 012033.

Bayuk, J.L., 2013. Security as a theoretical attribute construct. Comput. Secur. 37, 155–175.

Bozkus Kahyaoglu, S., Caliyurt, K., 2018. Cyber security assurance process from the internal audit perspective. Manage. Audit. J. 33 (4), 360–376.

Burton, J., 2023. Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. Technol. Soc. 102262.

Cheong, I., Caliskan, A., Kohno, T., 2022. Envisioning legal mitigations for LLM-based intentional and unintentional harms. Adm. Law J..

Cho, C.-S., Chung, W.-H., Kuo, S.-Y., 2015. Cyberphysical security and dependability analysis of digital control systems in nuclear power plants. IEEE Trans. Syst. Man Cybern. Syst. 46 (3), 356–369.

Darraj, E., Sample, C., Justice, C., 2019. Artificial intelligence cybersecurity framework: Preparing for the here and now with ai. In: ECCWS 2019 18th European Conference on Cyber Warfare and Security. Vol. 132, Academic Conferences and Publishing Limited.

Dedeke, A., 2017. Cybersecurity framework adoption: using capability levels for implementation tiers and profiles. IEEE Secur. Priv. 15 (5), 47–54.

Dhirani, L.L., Mukhtiar, N., Chowdhry, B.S., Newe, T., 2023. Ethical dilemmas and privacy issues in emerging technologies: a review. Sensors 23 (3), 1151.

Dykstra, J., Met, J., Backert, N., Mattie, R., Hough, D., 2022. Action bias and the two most dangerous words in cybersecurity incident response: An argument for more measured incident response. IEEE Secur. Priv. 20 (3), 102–106.

Ekambaranathan, A., Zhao, J., Van Kleek, M., 2023. How can we design privacy-friendly apps for children? Using a research through design process to understand developers' needs and challenges. Proc. ACM Hum.-Comput. Interact. 7 (CSCW2), 1–29.

Ekelund, S., Iskoujina, Z., 2019. Cybersecurity economics–balancing operational security spending. Inf. Technol. People 32 (5), 1318–1342.

Febriyani, W., Alhari, M.I., Kusumasari, T.F., 2022. Design of IT governance based on cobit 2019: A case study of XYZ education foundation. In: 2022 1st International Conference on Information System & Information Technology. ICISIT, IEEE, pp. 289–294.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al., 2021. An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. In: Ethics, Governance, and Policies in Artificial Intelligence. Springer, pp. 19–39.

Fujs, D., Mihelič, A., Vrhovec, S.L., 2019. The power of interpretation: Qualitative methods in cybersecurity research. In: Proceedings of the 14th International Conference on Availability, Reliability and Security. pp. 1–10.

Garvey, P.R., Patel, S.H., 2014. Analytical frameworks to assess the effectiveness and economic-returns of cybersecurity investments. In: 2014 IEEE Military Communications Conference. IEEE, pp. 136–145.

Goel, R., Kumar, A., Haddow, J., 2020. PRISM: a strategic decision framework for cybersecurity risk assessment. Inf. Comput. Secur. 28 (4), 591–625.

Gourisetti, S.N.G., Mylrea, M., Patangia, H., 2020. Cybersecurity vulnerability mitigation framework through empirical paradigm: Enhanced prioritized gap analysis. Future Gener. Comput. Syst. 105, 410–431.

Gourisetti, S.N.G., Mylrea, M., Reeve, H.M., Rotondo, J.A., Richards, G.T., Irwin, J.A., 2021. Facility Cybersecurity Framework Best Practices Version 2.0. Tech. Rep., Pacific Northwest National Lab. (PNNL), Richland, WA (United States).

Guha, N., Ho, D.E., Nyarko, J., Ré, C., 2022. Legalbench: Prototyping a collaborative benchmark for legal reasoning. arXiv preprint arXiv:2209.06120.

Gupta, M., Akiri, C., Aryal, K., Parker, E., Praharaj, L., 2023. From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. IEEE Access.

Hajny, J., Ricci, S., Piesarskas, E., Levillain, O., Galletta, L., De Nicola, R., 2021. Framework, tools and good practices for cybersecurity curricula. IEEE Access 9, 94723–94747.

Hitchcox, Z., 2020. Limitations of Cybersecurity Frameworks That Cybersecurity Specialists Must Understand to Reduce Cybersecurity Breaches (Ph.D. thesis). Colorado Technical University.

Hsu, C., Wang, T., Lu, A., 2016. The impact of ISO 27001 certification on firm performance. In: 2016 49th Hawaii International Conference on System Sciences. HICSS, IEEE, pp. 4842–4848.

Huang, H., Wu, S., Liang, X., Wang, B., Shi, Y., Wu, P., Yang, M., Zhao, T., 2023. Towards making the most of LLM for translation quality estimation. In: CCF International Conference on Natural Language Processing and Chinese Computing. Springer, pp. 375–386.

Iturbe, E., Rios, E., Rego, A., Toledo, N., 2023. Artificial intelligence for next generation cybersecurity: The AI4CYBER framework. In: Proceedings of the 18th International Conference on Availability, Reliability and Security. pp. 1–8.

Jarjoui, S., Murimi, R., 2021. A framework for enterprise cybersecurity risk management. In: Advances in Cybersecurity Management. Springer, pp. 139–161.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P., 2023. Survey of hallucination in natural language generation. ACM Comput. Surv. 55 (12), 1–38.

Kabanda, S., Tanner, M., Kent, C., 2018. Exploring SME cybersecurity practices in developing countries. J. Org. Comput. Electron. Commer. 28 (3), 269–282.

Karie, N.M., Sahri, N.M., Yang, W., Valli, C., Kebande, V.R., 2021. A review of security standards and frameworks for IoT-based smart environments. IEEE Access 9, 121975–121995.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al., 2023. ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Indiv. Differ. 103, 102274.

Katina, P.F., Keskin, O.F., 2021. Complex system governance as a foundation for enhancing the cybersecurity of cyber-physical systems. Int. J. Cyber Warfare Terror. (IJCWT) 11 (3), 1–14.

Kaur, R., Gabrijelčič, D., Klobučar, T., 2023. Artificial intelligence for cybersecurity: Literature review and future research directions. Inf. Fusion 101804.

Khader, M., Karam, M., Fares, H., 2021. Cybersecurity awareness framework for academia. Information 12 (10), 417.

Khan, F., Mer, A., 2023. Embracing artificial intelligence technology: Legal implications with special reference to European union initiatives of data protection. In: Digital Transformation, Strategic Resilience, Cyber Security and Risk Management. Emerald Publishing Limited, pp. 119–141.

King, Z.M., Henshel, D.S., Flora, L., Cains, M.G., Hoffman, B., Sample, C., 2018. Characterizing and measuring maliciousness for cybersecurity risk assessment. Front. Psychol. 9, 39.

Kissoon, T., 2020. Optimum spending on cybersecurity measures. Transform. Govern.: People Process Policy 14 (3), 417–431.

Kure, H.I., Islam, S., Mouratidis, H., 2022. An integrated cyber security risk management framework and risk predication for the critical infrastructure protection. Neural Comput. Appl. 34 (18), 15241–15271.

Leszczyna, R., 2021. Review of cybersecurity assessment methods: Applicability perspective. Comput. Secur. 108, 102376.

Li, L., He, W., Xu, L., Ash, I., Anwar, M., Yuan, X., 2019. Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. Int. J. Inf. Manage. 45, 13–24.

Li, Y., Liu, Q., 2021. A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments. Energy Rep. 7, 8176–8186.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., et al., 2023a. Summary of chatgpt-related research and perspective towards the future of large language models. Meta Radiol. 100017.

Liu, X., Tan, Y., Xiao, Z., Zhuge, J., Zhou, R., 2023b. Not the end of story: An evaluation of ChatGPT-driven vulnerability description mappings. In: Findings of the Association for Computational Linguistics: ACL 2023. pp. 3724–3731.

Maalem Lahcen, R.A., Caulkins, B., Mohapatra, R., Kumar, M., 2020. Review and insight on the behavioral aspects of cybersecurity. Cybersecurity 3 (1), 1–18.

Maglaras, L., Kantzavelou, I., Ferrag, M.A., 2021. Digital transformation and cybersecurity of critical infrastructures.

Malaivongs, S., Kiattisin, S., Chatjuthamard, P., 2022. Cyber trust index: A framework for rating and improving cybersecurity performance. Appl. Sci. 12 (21), 11174.

Malatji, M., Von Solms, S., Marnewick, A., 2019. Socio-technical systems cybersecurity framework. Inf. Comput. Secur. 27 (2), 233–272.

Manuel, D.-D., Carmona-Murillo, J., Cortés-Polo, D., Rodríguez-Pérez, F.J., 2022. CyberTOMP: A novel systematic framework to manage asset-focused cybersecurity from tactical and operational levels. IEEE Access 10, 122454–122485.

Markopoulou, D., Papakonstantinou, V., De Hert, P., 2019. The new EU cybersecurity framework: The NIS directive, ENISA's role and the general data protection regulation. Comput. Law Secur. Rev. 35 (6), 105336.

McIntosh, T., Liu, T., Susnjak, T., Alavizadeh, H., Ng, A., Nowrozy, R., Watters, P., 2023a. Harnessing GPT-4 for generation of cybersecurity GRC policies: A focus on ransomware attack mitigation. Comput. Secur. 134, 103424.

McIntosh, T.R., Liu, T., Susnjak, T., Watters, P., Ng, A., Halgamuge, M.N., 2023b. A culturally sensitive test to evaluate nuanced GPT hallucination. IEEE Trans. Artif. Intell. 1 (01), 1–13.

McIntosh, T.R., Susnjak, T., Liu, T., Watters, P., Halgamuge, M.N., 2024a. The inadequacy of reinforcement learning from human feedback - radicalizing large language models via semantic vulnerabilities. IEEE Trans. Cogn. Dev. Syst. 1 (01), 1–14.

McIntosh, T.R., Susnjak, T., Liu, T., Watters, P., Ng, A., Halgamuge, M.N., 2024b. A game-theoretic approach to containing artificial general intelligence: Insights from highly autonomous aggressive malware. IEEE Trans. Artif. Intell..

Meskó, B., Topol, E.J., 2023. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit. Med. 6 (1), 120.

Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heintz, I., Roth, D., 2023. Recent advances in natural language processing via large pre-trained language models: A survey. ACM Comput. Surv. 56 (2), 1–40.

Mirtsch, M., Kinne, J., Blind, K., 2020. Exploring the adoption of the international information security management system standard ISO/IEC 27001: a web mining-based analysis. IEEE Trans. Eng. Manage. 68 (1), 87–100.

Montagna, S., Ferretti, S., Klopfenstein, L.C., Florio, A., Pengo, M.F., 2023. Data decentralisation of LLM-based chatbot systems in chronic disease self-management. In: Proceedings of the 2023 ACM Conference on Information Technology for Social Good. pp. 205–212.

Nugraheni, D.M.K., Noranita, B., Adhy, S., Nugroho, A.K., 2022. Adopting COBIT 2019 for information technology risks in university online learning during COVID-19. In: COVID-19 Challenges to University Information Technology Governance. Springer, pp. 191–209.

Paskauskas, R.A., 2022. ENISA: 5G design and architecture of global mobile networks; threats, risks, vulnerabilities; cybersecurity considerations. Open Res. Eur. 2.

Pipyros, K., Thraskias, C., Mitrou, L., Gritzalis, D., Apostolopoulos, T., 2018. A new strategy for improving cyber-attacks evaluation in the context of tallinn manual. Comput. Secur. 74, 371–383.

Qi, X., Huang, K., Panda, A., Wang, M., Mittal, P., 2023. Visual adversarial examples jailbreak aligned large language models. In: The Second Workshop on New Frontiers in Adversarial Machine Learning.

Radanliev, P., De Roure, D., Nurse, J.R., Nicolescu, R., Huth, M., Cannady, S., Montalvo, R.M., 2018. Integration of cyber security frameworks, models and approaches for building design principles for the internet-of-things in industry 4.0. In: Living in the Internet of Things: Cybersecurity of the IoT-2018. IET, pp. 1–6.

Rathod, P., Hämäläinen, T., 2017. A novel model for cybersecurity economics and analysis. In: 2017 IEEE International Conference on Computer and Information Technology. CIT, IEEE, pp. 274–279.

Renaud, K., Ophoff, J., 2021. A cyber situational awareness model to predict the implementation of cyber security controls and precautions by SMEs. Organ. Cybersecur. J.: Pract. Process People 1 (1), 24–46.

Rjoub, G., Bentahar, J., Wahab, O.A., Mizouni, R., Song, A., Cohen, R., Otrok, H., Mourad, A., 2023. A survey on explainable artificial intelligence for cybersecurity. IEEE Trans. Netw. Serv. Manag..

Schuster, T., Fisch, A., Gupta, J., Dehghani, M., Bahri, D., Tran, V., Tay, Y., Metzler, D., 2022. Confident adaptive language modeling. Adv. Neural Inf. Process. Syst. 35, 17456–17472.

Shim, J.P., Sharda, R., French, A.M., Syler, R.A., Patten, K.P., 2020. The internet of things: Multi-faceted research perspectives. Commun. Assoc. Inf. Syst. 46 (1), 21.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., et al., 2023. Large language models encode clinical knowledge. Nature 620 (7972), 172–180.

Slapničar, S., Vuko, T., Čular, M., Drašček, M., 2022. Effectiveness of cybersecurity audit. Int. J. Account. Inf. Syst. 44, 100548.

Sule, M.-J., Zennaro, M., Thomas, G., 2021. Cybersecurity through the lens of digital identity and data protection: issues and trends. Technol. Soc. 67, 101734.

Sulistyowati, D., Handayani, F., Suryanto, Y., 2020. Comparative analysis and design of cybersecurity maturity assessment methodology using nist csf, cobit, iso/iec 27002 and pci dss. JOIV: Int. J. Inform. Vis. 4 (4), 225–230.

Syafrizal, M., Selamat, S.R., Zakaria, N.A., 2020. Analysis of cybersecurity standard and framework components. Int. J. Commun. Netw. Inf. Secur. 12 (3), 417–432.

Szabó, Z., Bilicki, V., 2023. A new approach to web application security: Utilizing GPT language models for source code inspection. Future Internet 15 (10), 326.

Taherdoost, H., 2022. Understanding cybersecurity frameworks and information security standards—a review and comprehensive overview. Electronics 11 (14), 2181.

Tawalbeh, L., Muheidat, F., Tawalbeh, M., Quwaider, M., 2020. IoT privacy and security: Challenges and solutions. Appl. Sci. 10 (12), 4102.

Tissir, N., El Kafhali, S., Aboutabit, N., 2021. Cybersecurity management in cloud computing: semantic literature review and conceptual framework proposal. J. Reliable Intell. Environ. 7, 69–84.

Toufiq, M., Rinchai, D., Bettacchioli, E., Kabeer, B.S.A., Khan, T., Subba, B., White, O., Yurieva, M., George, J., Jourde-Chiche, N., et al., 2023. Harnessing large language models (LLMs) for candidate gene prioritization and selection. J. Transl. Med. 21 (1), 728.

Triplett, W.J., 2022. Addressing human factors in cybersecurity leadership. J. Cybersecur. Priv. 2 (3), 573–586.

Tvaronavičienė, M., Plėta, T., Della Casa, S., Latvys, J., 2020. Cyber security management of critical energy infrastructure in national cybersecurity strategies: Cases of USA, UK, France, Estonia and Lithuania. Insights Reg. Dev. 2 (4), 802–813.

Ukil, A., Gama, J., Jara, A.J., Marin, L., 2023. Knowledge-driven analytics and systems impacting human quality of life-neurosymbolic AI, explainable AI and beyond. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. pp. 5296–5299.

Wang, Z., Xie, W., Chen, K., Wang, B., Gui, Z., Wang, E., 2023. Self-deception: Reverse penetrating the semantic firewall of large language models. arXiv preprint arXiv:2308.11521.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., et al., 2021. Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.

Winograd, A., 2023. Loose-lipped large language models spill your secrets: the privacy implications of large language models. Harvard J. Law Technol. 36 (2).

Yang, Z., Li, L., Lin, K., Wang, J., Lin, C.-C., Liu, Z., Wang, L., 2023. The dawn of LMMs: Preliminary explorations with GPT-4V (ision). arXiv preprint arXiv:2309.17421. Affiliation: Microsoft.

Yeoh, W., Wang, S., Popovič, A., Chowdhury, N.H., 2022. A systematic synthesis of critical success factors for cybersecurity. Comput. Secur. 118, 102724.

Yusif, S., Hafeez-Baig, A., 2021. A conceptual model for cybersecurity governance. J. Appl. Secur. Res. 16 (4), 490–513.

Zhang, P., Kamel Boulos, M.N., 2023. Generative AI in medicine and healthcare: Promises, opportunities and challenges. Future Internet 15 (9), 286.

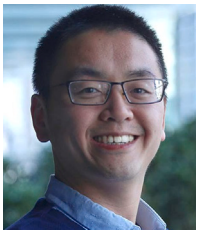
Zhang, H., Song, H., Li, S., Zhou, M., Song, D., 2022. A survey of controllable text generation using transformer-based pre-trained language models. ACM Comput. Surv..

Dr. Timothy R. McIntosh received his Ph.D. in cybersecurity from La Trobe University (Melbourne) in 2022. He is currently the Generative AI & Cybersecurity Research Strategist at Cyberoo Pty Ltd, Surrey Hills (Australia), and an adjunct lecturer in cybersecurity at La Trobe University, Melbourne (Australia). His research focuses on LLMs in cybersecurity, and ransomware mitigation. He is an ISC2 CISSP and CSSLP, an IAPP CIPP/E and CIPT, a CompTIA CASP+ and CySA+, a Microsoft Certified Cybersecurity Architect Expert, and a Microsoft MCSD.

Dr. Teo Susnjak is a Senior Lecturer in Computer Science and Information Technology at Massey University (Auckland, New Zealand), where he received his Ph.D. in machine learning in 2011. He is the Subject Lead of Data Science and the Coordinator for the Master of Analytics degree at Massey University. His recent research focused on *Large Language Models* (LLMs) for tasks such as generating data visualizations, fine-tuning data visualizations, sentiment analysis of scientific literature, and exploring the implications of generative AI technologies in Education.

Dr. Tong Liu is an accomplished computer science and information technology lecturer at Massey University's School of Mathematical and Computational Sciences in Auckland, New Zealand. Her research is primarily focused on machine learning and artificial intelligence, and she has developed several algorithms in this field. Dr. Liu's work has received multiple research grants, and she has also been a guest editor for a special issue of the *Pattern Recognition Letter* journal.

Prof. Paul Watters serves as the CEO of Cyberstronomy Pty Ltd, (Ballarat). He has previously held notable academic positions, being the Academic Dean at Academies Australasia Polytechnic, Melbourne (Australia), an Honorary Professor at Macquarie University, Sydney(Australia), and an Adjunct Professor at La Trobe University. He is a Chartered IT Professional, a Fellow of the British Computer Society, a Senior Member of the IEEE, a Member of the ACM, and a Member of the Australian Psychological Society.



Dan Xu obtained his Master of Applied Science (Information Systems) from RMIT University (Melbourne) in 2008, and Master of Computing from Federation University (Ballarat) in 2017. He currently works as a CyberCrime Manager at ANZ Bank, specializing in scam analysis. His role involves investigating and mitigating various types of cyber scams, including phishing, advance fee fraud, and social engineering attacks. He is a CISSP, and also has a strong background in malware reverse engineering, which has equipped him with a deep understanding of the technical aspects of cyber threats.

Dongwei Liu is an accomplished enterprise systems analyst with a Master of Applied Science in Information Systems and over 10 years of industry experience. He excels in database warehousing, SQL reporting, systems analysis, business intelligence, and systems development. Dongwei specializes in UKG (Kronos), Chris21, Outsystems, UiPath, Microsoft Business Intelligence, and SharePoint. He was awarded the Ramsay Health Care Excellence in Leadership in 2019 and is currently a Senior Software Engineer at Coles Group.

Raza Nowrozy is currently pursuing his Ph.D. in cybersecurity at Victoria University in Melbourne, Australia. He obtained his Bachelor of Computer and Information Science from the University of South Australia back in 2004, and completed his Master of Cybersecurity from La Trobe University in Melbourne in 2019. He holds a certification as an ISACA CISM and serves as a tech lead in the cybersecurity industry. Raza's research focuses on the area of Health Information, with a specific interest in the privacy and security of My Health Record (MHR).

Dr. Malka N. Halgamuge obtained her Ph.D. from the University of Melbourne (Melbourne) in 2007, and worked there as a researcher during 2007–2021. She was awarded prestigious fellowships at the University of California, Los Angeles (USA), Lund University, Lund (Sweden), and the Chinese Academy of Sciences (CAS) in Beijing (PRC), among others. She is currently a Senior Lecturer in cybersecurity at RMIT University, Melbourne, Australia. Her research includes emerging technology, cybersecurity, security in IoT, edge, power grid, supply chain, blockchain, and developing solutions using deep/machine learning.