Towards Securing AI Integrated Business Processes: a Reasoning Framework Informed by a Semantic Model

Gal Engelberg[1,†], Avi Shaked[2,*,†], Nan Messe[3,†] and Pnina Soffer[1,†]

[1]University of Haifa, Haifa, Israel
[2]Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel
[3]IRIT, Toulouse University, CNRS, INP, UT2, Toulouse, France

Abstract

The integration of large language models (LLMs) and Artificial Intelligence (AI) into business processes is transforming organizational workflows, by automating tasks of increasing complexity. This incorporation of agentic AI into business processes introduces new security risks that are often overlooked at the business level. Existing frameworks focus on application-level controls, leaving a gap in understanding how AI-related vulnerabilities impact overall process behavior. This preliminary work suggests a semantic-informed framework to support organizations and individuals in assessing security implications of AI business process transformations. The visionary framework relies on a semantic model to link business process models with agentic AI components, capturing key security objectives, threats, mitigation strategies, and their interdependencies; and will use an reasoning mechanism to inform users of potential violations of security objectives when integrating AI into business processes and assist them in re-designing the processes to account for the potential violations. We demonstrate the suggested framework for a recruitment business process.

Keywords

AI-Augmented Workflows, Business Process Security, Conceptual Modeling, AI Security, Responsible BPM

1. Introduction

We are in an Artificial Intelligence (AI) driven transformation era, as organizations rapidly automate core business processes using recent AI advancements [1, 2]. Technology providers and enterprises embed generative AI into platforms to deploy autonomous agents that streamline workflows, handle queries, improve efficiency, and user satisfaction [3, 4]. Academic and industry research supports this shift, showing how AI can enhance business process management (BPM) [5, 6, 7]. However, this may introduce new security risks.

BPM-related security has traditionally focused on infrastructure risks from cloud platforms, Information Technology (IT), and Internet of Things (IoT). Cloud-based BPM poses confidentiality and integrity risks due to multi-tenancy and shared services [8]. IoT integration exposes processes to edge device vulnerabilities, potentially disrupting availability [9]. In IT environments, cyberattacks propagate through interdependent systems, impacting business continuity [10]. Existing work often assumes deterministic systems, overlooking emergent risks posed by AI-driven decision-making.

AI agents are autonomous systems that combine Large Language Models (LLM) based language processing with tool integration to perform complex tasks without human input, enabling human-like reasoning and communication across domains [11]. Integrating AI agents into business processes introduces unique security risks absent in traditional systems. For example, in recruitment processes that incorporate AI systems, prompt injection attacks may embed hidden instructions in resumes to manipulate evaluations or extract sensitive data as hiring criteria [12]. This example reflects a broader

RCIS 2026: Companion Proceedings of the 20th Conference on Research Challenges in Information Science: RCIS Research Projects and Workshops, May 26-29, 2026, Toulouse, France

*Corresponding author.
†These authors contributed equally.
gal.engelberg@gmail.com (G. Engelberg); avishaked@tauex.tau.ac.il (A. Shaked); nan.messe@irit.fr (N. Messe); spnina@is.haifa.ac.il (P. Soffer)

set of challenges related to the reliability, transparency, and governability of AI agent behavior in business contexts [6, 7].

The AI transformation has led to new security frameworks and standards. Examples include OWASP Top 10 for LLM Applications [12], outlining threats such as prompt injection and insecure output handling; OWASP Agentic AI [13], addressing threats arising from AI autonomy; MITRE Atlas [14], cataloging adversarial techniques targeting AI systems; and NIST AI Risk Management Framework [15], offering structured governance of AI risks across the trustworthiness and accountability dimensions. These frameworks provide taxonomies of key threats and control mechanisms in the agentic AI landscape. Also, while well-established security knowledge bases – such as MITRE's Common Attack Pattern Enumeration and Classification (CAPEC) [16] – remain relevant, they typically require some interpretation to translate into AI context.

Regulations, such as the EU AI Act [17], aim to ensure safe and responsible AI use. They introduce risk-based classifications, imposing stricter requirements on higher risk systems. From a security perspective, the EU AI Act emphasizes resilience to manipulation, reliable data management, transparency of system behavior, and mechanisms for human oversight. Applied to business processes, it underscores the need for careful AI integration, particularly where automated decision affects employees, customers, or regulatory compliance. Specifically, the use of AI systems in recruitment processes is considered – by the EU AI act – as high-risk, since *"those systems may appreciably impact future career prospects and livelihoods of... persons"*, *"perpetuate historical patterns of discrimination"*, and *"impact... rights to data protection and privacy"*.

The various security frameworks address risks at the application system level [18, 19], leaving the business process level underexplored. This creates gaps in understanding the full business impact of AI-related vulnerabilities. There is an inability to reason about how AI-specific vulnerabilities impact end-to-end business processes and their security properties. This limits the analysis of technical risks and their mitigation – as outlined in various security frameworks – with respect to goal-oriented regulatory requirements – such as those introduced by the AI Act – and to business objectives. To address this, we propose a framework that provides foundations for reasoning about the security impact of integrating AI agents into business processes. In this design science research, we investigate how a semantic model and reasoning mechanisms can support the identification and mitigation of security threats arising from the integration of agentic AI into business processes.

The paper is structured as follows: Section 2 introduces the framework and the methodology. Section 3 presents preliminary results demonstrating how the semantic model supports reasoning and decision-making for integrating AI agents into a recruitment process as a representative business processes. Last, conclusion is provided, in Section 4.

2. Semantic Model Informed Reasoning Framework

As a first step towards addressing the outlined challenges, we propose a framework for reasoning about the security impacts of integrating AI agents into business processes. The framework, sketched in Figure 1, consists of two main components. The first is a semantic model that codifies descriptions of business processes, AI agents, threats and properties relevant to agentic AI technologies in the context of business processes. The second component is an inference mechanism to reason about the potential effects of integrating agentic AI into the business process.

Our research methodology consists of three core steps: (1) identifying relevant concepts from the literature and mapping their dependencies; (2) codifying business processes and their constituent activities alongside their desirable security properties; and (3) analyzing how AI agents may compromise the desirable business process security properties. We exercise these steps in the next section, where we provide a basic, high-level example.

We plan to evaluate the semantic model using security-related domain ontologies [20, 21], and to assess the usefulness of the inference through interviews with domain experts. We are also working towards the implementation of a modelling environment that would allow to model process design

scenarios and reason about agentic AI integration from a security perspective.

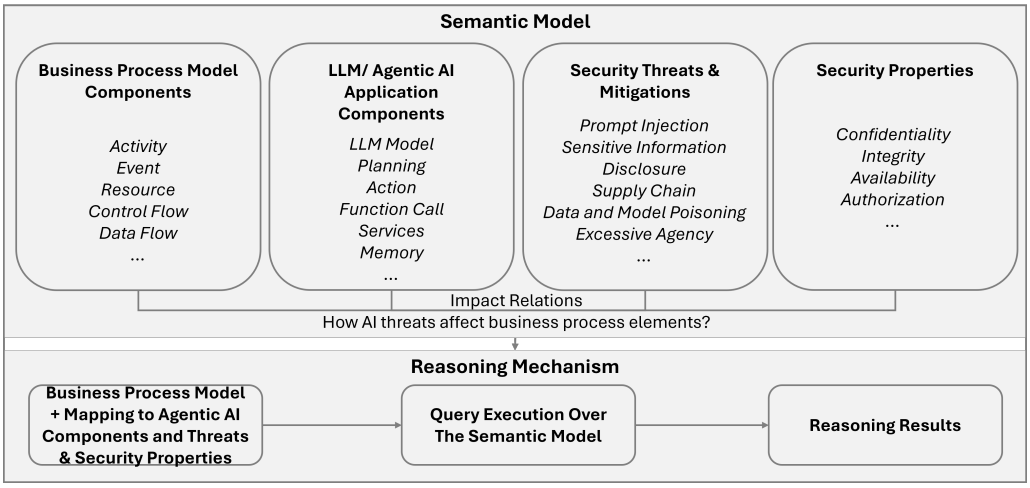

Figure 1: The Proposed Semantic Model Informed Reasoning Framework

3. Demonstrating Reasoning about AI Security Threats in Business Processes

We illustrate how the reasoning framework – informed by a semantic model – can support decision-making with respect to incorporating AI agents in a recruitment process, which is a prominent business process in organizations across virtually every domain. Pertinent threats include prompt injection attacks, where a malicious user embeds hidden instructions within their resume, manipulating the AI to provide favorable evaluations irrespective of actual qualifications; and extracting sensitive company information, including confidential hiring criteria or other protected data [12].

We exemplify our approach using Object-Process Methodology (OPM) [22] models. OPM was selected for its sound underlying ontology as well as its inherent ability to capture stateful objects, processes, their relations and hierarchies as key design elements. Also, we found that OPM's simple visual language (Object Process Diagram) and the corresponding, automatically generated natural language description of the model (Object Process Language) are useful in illustrating the envisioned framework, including both the semantic model and its reasoning results. However, additional process and reasoning representations – such as Business Process Model and Notation (BPMN) and Prolog – will be considered when we design the full decision support system.

Our preliminary modelling effort is available at https://ash-systems.github.io/AI-BPM/OPM/BPM_AI.html. Note that the semantic model, which is an ongoing effort, is not shown here, but it informs the elements included in the demonstrative OPM model of the selected recruitment process. Fig. 2 shows an excerpt from our model, developed in two phases. The top part presents the business process and its security requirements, resulting from exercising step #2 of the methodology. The bottom part introduces agentic AI components and their associated threats, as a result of exercising step #1 of the methodology, as well as some relations inferred by employing step #3 of the methodology (in pink). The automatically generated natural language model representation of the model (using Object Process Language) appears in Fig. 3.

As described in the top model, the *Recruiting For A Position* business process includes two specific activities: *Receiving Applications* and *Selecting Top Applications*. A third activity is listed simply to denote that the process may include other activities not explicitly discussed here. On the left side, we list relevant security properties that are related to the activities. *Response Integrity* is a specific form of the known *Integrity* property, and *Intellectual Property Confidentiality* is a specific form of the *Confidentiality* property. We note that while the *Intellectual Property Confidentiality* property often

appears in non AI security properties taxonomies (e.g., [23]), the *Response Integrity* property is typically absent, and can be seen as an AI-specific flavor of *Resource trustfulness*. A property has two states: *ok*, signifying it is satisfied, and *not ok* if it has the potential to be violated. The security requirements of the process could be determined by assigning properties to activities. For instance, *Receiving Applications* is assigned the *Intellectual Property Confidentiality* property, since it is facing external candidates; and *Selecting Top Applications* is assigned the *Response Integrity* property, signifying correctness of the selection process. On the right side of the model, the process resources – Specific Human Role and Specific AI Agent – are listed.

The bottom model includes potential specializations, to the level of specific agents. Some relevant components of the *Specific AI Agent* appear: *LLM Model*, *Function Calling Element*, *Memory Element*. TheseIn addition, we outlined threats that are relevant to our example. We are working on integrating an exhaustive list of such threats into the semantic model. Here, we discuss two threats adopted from OWASP Top 10 for LLM Applications [12]: *Prompt Injection* (LLM01:2025); and Sensitive Information Disclosure (LLM02:2025). By analysis we establish these threats as pertinent to the LLM Model component (of the AI agent). Accordingly, they are linked with their potential adversarial effect of violating specific properties. Two other threats appear, relating to the other AI agent components. While these will not be further discussed, we mention that one of them – *Embedding Extraction* – is a subtype of another OWASP Top 10 for LLM Applications threat (LLM08:2025 *Vector and Embedding Weaknesses*); and the other – *Tool Misuse* – was recently identified as *the ASI02: Tool Misuse and Exploitation* threat in OWASP Top 10 For Agentic Applications 2026 [24], which was released after we completed our modelling effort. It is also strongly related to the CAPEC threat *Functionality Misuse* (CAPEC-212) [16].

The component-threat-property mappings allow for a reasoning mechanism, as outlined in Fig. 1, to determine what are the properties that a specific AI agent exhibits, with respect to both compliance/satisfaction (*ok* states) and violation (*not ok* states). These are represented by purple relations in Fig. 2, indicating they emerge from other model elements by using the reasoning mechanism. These derived relations also appear in line 13 in Fig. 3.

An example of a reasoning query over the presented semantic model, could be: *Is an AI agent of the specific AI agent type suitable for performing both activities involved in recruiting for a position?* The reasoning mechanism would evaluate if any property violations associated with the AI agent conflicts with the required properties of each activity. The results of such analysis could be communicated with the process owner/designer so that – if needed – a different agent would be used and/or a mitigation could be incorporated. For example, since a *Prompt Injection* by candidates could lead to the *Selecting Top Applications* preferring their applications (over others that are more suitable for the position), the desirable *Response Integrity* of this activity may be compromised if handled exclusively by the *Specific AI Agent*. A proper mitigation could be in the form of involving a human agent as the co-handler of this activity (e.g., performing lower-level inspection and results verification activities).

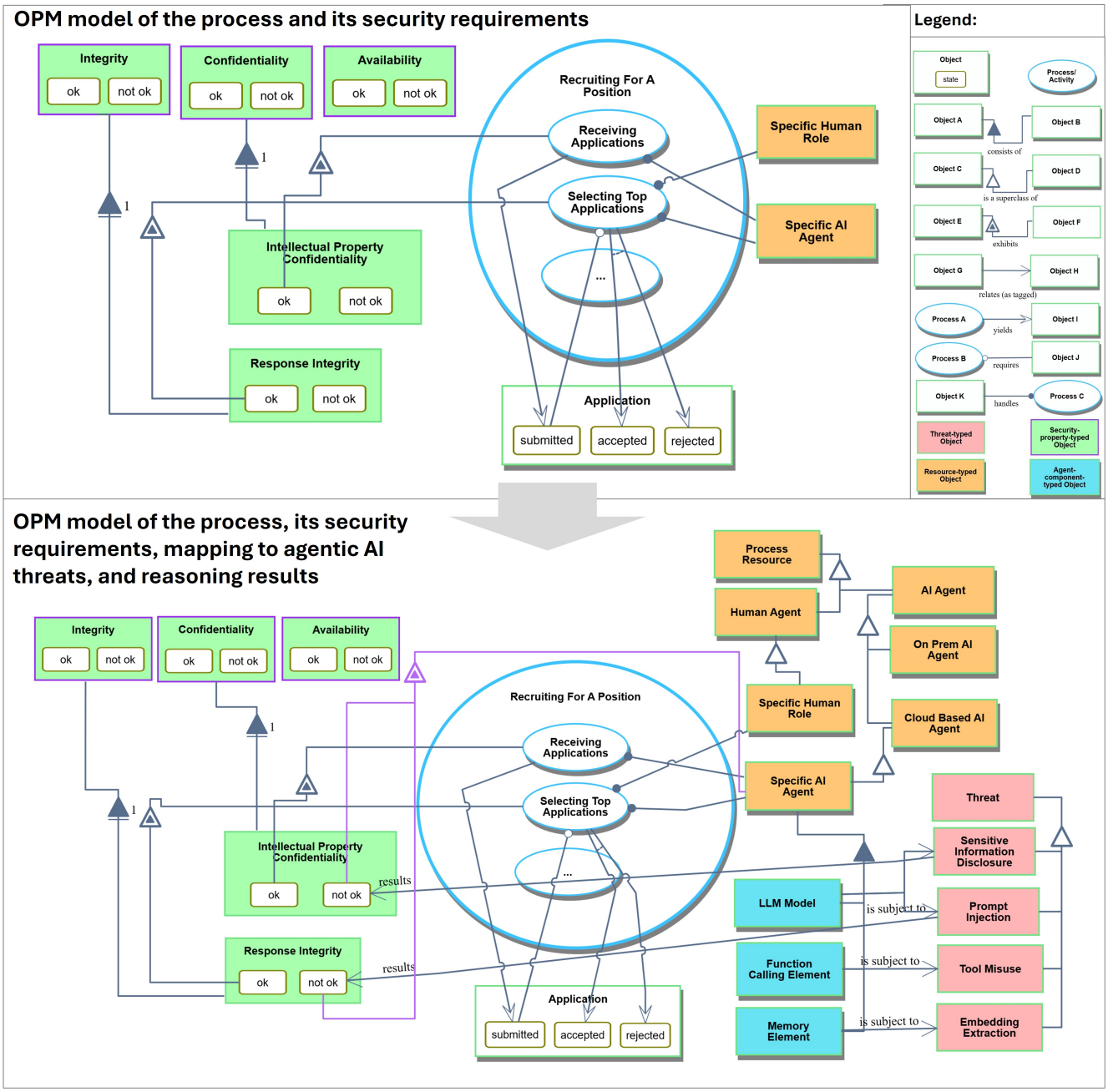

Figure 2: Diagrammatic representation (Object Process Diagram) of the example recruitment process

OPL

1. Recruiting For A Position from SD zooms in SD1 into Receiving Applications, Selecting Top Applications, and ..., which occur in that time sequence.
2. Intellectual Property Confidentiality can be not ok or ok.
3. Response Integrity can be not ok or ok.
4. Availability can be not ok or ok.
5. Confidentiality can be not ok or ok.
6. Integrity can be not ok or ok.
7. Application can be accepted, rejected or submitted.
8. AI Agent and Human Agent are Process Resources.
9. Cloud Based AI Agent and On Prem AI Agent are AI Agents.
10. Confidentiality consists of Intellectual Property Confidentiality and five more parts.
11. Specific AI Agent consists of Function Calling Element, LLM Model, and Memory Element.
12. Specific AI Agent is a Cloud Based AI Agent.
13. Specific AI Agent exhibits Intellectual Property Confidentiality with value not ok and Response Integrity with value not ok.
14. Embedding Extraction, Prompt Injection, Sensitive Information Disclosure, and Tool Misuse and two more specializations are Threats.
15. LLM Model is subject to Prompt Injection and Sensitive Information Disclosure.
16. Sensitive Information Disclosure results state not ok of Intellectual Property Confidentiality.
17. Prompt Injection results state not ok of Response Integrity.
18. Selecting Top Applications exhibits Response Integrity with value ok.
19. Receiving Applications exhibits Intellectual Property Confidentiality with value ok.
20. Integrity consists of Response Integrity and 17 more parts.
21. Memory Element is subject to Embedding Extraction.
22. Function Calling Element is subject to Tool Misuse.
23. Specific Human Role is a Human Agent.
24. Specific AI Agent handles Receiving Applications.
25. Receiving Applications yields Application at state submitted.
26. Specific AI Agent and Specific Human Role handle Selecting Top Applications.
27. Selecting Top Applications requires Application at state submitted.
28. Selecting Top Applications yields Application at one of the states accepted or rejected.

Figure 3: Textual, natural language representation (Object Process Language) of the example recruitment process

4. Conclusion

The introduction of AI agents into business processes may compromise desirable security properties. We propose a semantic-informed reasoning framework to evaluate such introduction and assist in its design for security. Our framework relies on a semantic model that captures key concepts – such as AI agent components and their constituents, business process constituents (activities, flows and states) – and their relations, as well as a reasoning mechanism.

We exemplified our reasoning framework using the high-risk scenario of a recruitment process. Our demonstration illustrates the value of our framework in integrating security definitions from multiple bodies of knowledge and providing actionable insights. While our example uses OPM modelling, the leading OPM platform (OPCloud) provides limited support for reasoning. We intend to examine other modelling alternatives that allow for better integration with descriptive logic reasoners and analysis capabilities (e.g., domain specific modelling [25] and SysML v2 [26]).

We envision the framework as a decision support system, to inform decisions of using AI agents, by individuals as well as organizations. By better framing the contribution of integrating AI to the business processes and understanding its implications, we expect individuals and organizations to use AI more responsibly. We are further developing and refining the underlying semantic model. Specifically, we aim to incorporate exhaustive catalogs of agentic AI threats, components, business process security properties and their relations. We also plan to include mitigation patterns, to support users in properly transforming their business processes into AI-integrated processes. Furthermore, we intend to provide a formal description of the proposed reasoning mechanism and implement an automated version that would allow to verify business process models with respect to AI agents integration, to support decision-makers and process designers.

Acknowledgments

This work was supported by the Center for Cyber Law & Policy at the University of Haifa in conjunction with the Israel National Cyber Directorate in the Prime Minister's Office.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Gemini and Copilot for phrasing and proofreading. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] E. Chrzanowska, M. Chrzanowski, P. Zawada, Ai-powered digital transformation–organizational perspective. literature review, Journal of Modern Science 60 (2024) 429–442.

[2] M. Wornow, A. Narayan, K. Opsahl-Ong, Q. McIntyre, N. Shah, C. Ré, Automating the enterprise with foundation models, Proceedings of the VLDB Endowment 17 (2024) 2805–2812.

[3] ServiceNow, AI agents studio, 2025. URL: https://www.servicenow.com/docs/r/intelligent-experiences/ai-agent-studio.html, accessed: 2026-04-27.

[4] IBM, Ibm watsonx orchestrate, 2026. URL: https://www.ibm.com/products/watsonx-orchestrate, accessed: 2026-04-27.

[5] M. Dumas, F. Fournier, L. Limonad, A. Marrella, M. Montali, J.-R. Rehse, R. Accorsi, D. Calvanese, G. De Giacomo, D. Fahland, et al., Ai-augmented business process management systems: a research manifesto, ACM Transactions on Management Information Systems 14 (2023) 1–19.

[6] M. Vidgof, S. Bachhofner, J. Mendling, Large language models for business process management: Opportunities and challenges, in: International conference on business process management, Springer, 2023, pp. 107–123.

[7] H. Vu, N. Klievtsova, H. Leopold, S. Rinderle-Ma, T. Kampik, Agentic business process management: the past 30 years and practitioners' future perspectives, arXiv e-prints (2025) arXiv–2504.
[8] N. Soveizi, F. Turkmen, D. Karastoyanova, Security and privacy concerns in cloud-based scientific and business workflows: A systematic review, Future Generation Computer Systems 148 (2023) 184–200.
[9] M. Hornsteiner, L. Kölbel, D. Oberhofer, S. Schönig, A reflection on process-oriented industrial iot security management., in: ICISSP 2025), 2025, pp. 242–253.
[10] O. González-Rojas, N. Castro, S. Lesmes, Quantifying risk propagation within a network of business processes and it services: O. gonzález-rojas et al., Business & Information Systems Engineering 63 (2021) 129–143.
[11] OpenAI, A practical guide to building agents, https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf, 2024. Accessed: 2026-04-27.
[12] OWASP Foundation, OWASP Top 10 for Large Language Model Applications, https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-v2025.pdf, 2025. Version 1.1, Accessed: 2026-04-27.
[13] OWASP GenAI Security Project, Agentic AI – threats and mitigations, Online resource, OWASP GenAI Security Project, 2025. URL: https://genai.owasp.org/resource/agentic-ai-threats-and-mitigations/, accessed: 2026-04-27.
[14] MITRE Corporation, MITRE ATLAS, https://atlas.mitre.org/, 2024. Accessed: 2026-04-27.
[15] E. Tabassi, AI RMF 1.0, Technical Report NIST AI 100-1, NIST, Gaithersburg, MD, 2023. URL: https://doi.org/10.6028/NIST.AI.100-1, accessed: 2026-04-27.
[16] MITRE Corporation, MITRE CAPEC, https://capec.mitre.org/, 2024. Accessed: 2026-04-27.
[17] European Commission, Eu artificial intelligence act, 2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, accessed: 2026-04-27.
[18] L. Wu, C. Wang, T. Liu, Y. Zhao, H. Wang, From assistants to adversaries: Exploring the security risks of mobile llm agents, 2025.
[19] A. Shaked, N. Messe, Bridgesec: Facilitating effective communication between security engineering and systems engineering, Journal of information security and applications 89 (2025) 103954.
[20] Í. Oliveira, T. P. Sales, R. Baratella, M. Fumagalli, G. Guizzardi, An ontology of security from a risk treatment perspective, in: International conference on conceptual modeling, Springer, 2022, pp. 365–379.
[21] T. P. Sales, F. Baião, G. Guizzardi, J. P. A. Almeida, N. Guarino, J. Mylopoulos, The common ontology of value and risk, in: International conference on conceptual modeling, Springer, 2018, pp. 121–135.
[22] D. Dori, Object-Process Methodology: A Holistic Systems Paradigm, Springer, 2002.
[23] I. Sayar, N. Messe, S. Ebersold, J.-M. Bruel, From what to how: A taxonomy of formalized security properties, 2025.
[24] OWASP Foundation, OWASP Top 10 For Agentic Applications 2026, https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/, 2025. Version 2026, Accessed: 2026-04-27.
[25] A. Shaked, N. Messe, T. Melham, Modelling tool extension for vulnerability management, in: Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems, 2024, pp. 56–60.
[26] A. Ahlbrecht, W. Zaeske, U. Durak, Mbsqle: Enabling sql as powerful query language for SysML v2 in aviation, in: 2025 AIAA DATC/IEEE 44th Digital Avionics Systems Conference (DASC), IEEE, 2025, pp. 1–10.