Universidade Federal da Paraíba
Centro de Ciências Humanas, Letras e Artes
Departamento de Filosofia
Curso de Pós-Graduação Lato Sensu em
Ética em Inteligência Artificial, EaD

Roberta Brandão do Nascimento

An Ontology for Ethical Risk Analysis in AI-Enabled SOAR Systems.

João Pessoa
2026

Roberta Brandão do Nascimento

An Ontology for Ethical Risk Analysis in AI-Enabled SOAR Systems.

Trabalho de conclusão de curso apresentado ao Curso de Pós-Graduação Lato Sensu em Ética em Inteligência Artificial, EaD, do Departamento de Filosofia, do Centro de Ciências Humanas, Letras e Artes, da Universidade Federal da Paraíba, como requisito para obtenção do título de Especialista em Ética em Inteligência Artificial

Orientador: Dr. Patrício de Alencar Silva

João Pessoa
2026

Catalogação na publicação
Seção de Catalogação e Classificação

N244o Nascimento, Roberta Brandão do.
An Ontology for Ethical Risk Analysis in AI-Enabled SOAR Systems / Roberta Brandão do Nascimento. - João Pessoa, 2026.
13 f. : il.

Orientador : Patrício de Alencar Silva.
TCC (Especialização) - Universidade Federal da Paraíba/Centro de Ciências Humanas, Letras e Artes, 2026.

1. Cybersecurity governance. 2. Ethical risks. 3. AI-based SOAR. 4. Ontology. 5. OntoUML. I. Silva, Patrício de Alencar. II. Título.

UFPB/CCHLA CDU 004.8

Elaborado por KARLA MARIA DE OLIVEIRA - CRB-15/485

FOLHA DE APROVAÇÃO

Roberta Brandão do Nascimento

An Ontology for Ethical Risk Analysis in AI-Enabled SOAR Systems

Trabalho de conclusão de curso apresentado ao Curso de Pós-Graduação *Lato Sensu* em Ética em Inteligência Artificial, EaD, do Departamento de Filosofia, do Centro de Ciências Humanas, Letras e Artes, da Universidade Federal da Paraíba, como requisito para obtenção do título de Especialista em Ética em Inteligência Artificial.

Aprovado em: 30/06/2026.

Banca Examinadora

_____

Prof. Doutor Patrício de Alencar Silva
Orientador
(UFPB/CCHLA/DEPARTAMENTO DE FILOSOFIA)

_____

Prof. Doutor Danilo Fraga Dantas
Membro Interno
(UFPB/CCHLA/DEPARTAMENTO DE FILOSOFIA)

_____

Prof. Doutor Lucídio dos Anjos Formiga Cabral
Membro Interno
(UFPB/CCHLA/DEPARTAMENTO DE FILOSOFIA)

An Ontology for Ethical Risk Analysis in AI-Enabled SOAR Systems

Roberta Brandão do Nascimento[1], Patrício de Alencar Silva[2]

[1]Universidade Federal da Paraíba, Cidade Universitária, Castelo Branco, João Pessoa - PB, 58051-900, Brazil
[2]Universidade Federal Rural do Semi-Árido, Av. Francisco Mota, 572, Bairro Costa e Silva, Mossoró - RN, 59625-900, Brazil

Abstract

Security Orchestration, Automation, and Response (SOAR) platforms have become increasingly important in cybersecurity operations by integrating heterogeneous security tools, automating repetitive tasks, and coordinating incident response activities. More recently, Artificial Intelligence (AI) techniques have been incorporated into SOAR environments to improve alert triage, anomaly detection, prioritization, and response execution. Although this integration enhances operational scalability and analytical capacity, it also raises ethical issues in the development and use of this technology combination. Hence, a relevant research question arising from this combination is: *what ethical issues arise from combining AI with SOAR, and how they are represented?* To address this problem, this paper presents an ontology to represent ethical issues arising from the combination of AI with SOAR. The ontology is grounded in an Ontology Requirements Specification Document (ORSD) and modeled in OntoUML, a graphical ontology modeling language based on the Unified Foundational Ontology (UFO). The proposed *task ontology* organizes the domain's structural and dynamic aspects into four views: *technical infrastructure, operational processes, governance and supervision*, and *ethical consequences and risks*. The state-of-the-art tools for OntoUML modeling support semi-automated ontology syntax verification and promote semantic interoperability by serializing the model into Semantic Web languages such as JSON and gUFO. The ontology model is evaluated through a critical analysis grounded in the principles of ontological modeling in OntoUML and the technical guidelines of Cybersecurity governance.

Keywords

Cybersecurity Governance, Ethical Risks, AI-based SOAR, Ontology, OntoUML

1. Introduction

The increasing complexity, scale, and velocity of cyber threats have significantly transformed the operational demands imposed on contemporary Security Operations Centers (SOC). Organizations must continuously monitor heterogeneous digital infrastructures while processing large volumes of alerts generated by security information and event management systems, intrusion detection systems, endpoint monitoring tools, and threat intelligence platforms. In this context, cybersecurity operations increasingly depend on technological mechanisms that integrate distributed sources of information, reduce manual effort, and accelerate incident response [1].

Security Orchestration, Automation, and Response (SOAR) platforms have emerged as an important architectural component of this operational landscape. SOAR systems support the integration of heterogeneous security technologies, automate repetitive tasks, and coordinate incident response activities through predefined operational workflows commonly referred to as playbooks. These platforms improve operational efficiency, reduce response times, and increase the scalability of security operations [1].

More recently, Artificial Intelligence (AI) techniques have been progressively incorporated into SOAR environments [2]. Machine learning models and data-driven analytical mechanisms are increasingly used to support alert classification, anomaly detection, prioritization of security events, and automated or semi-automated response recommendations [3]. In some operational contexts, AI-enabled SOAR systems may autonomously trigger response actions based on inferred classifications or predictive

Proceedings of the 19th Seminar on Ontology Research in Brazil (ONTOBRAS 2026) and 10th Doctoral and Masters Consortium on Ontologies (WTDO 2026), Vitória (ES), Brazil, September 23, 2026.
Full Article
betabrandao@gmail.com (R. B. d. Nascimento); patricio.alencar@ufersa.edu.br (P. d. A. Silva)
0009-0003-2060-9470 (R. B. d. Nascimento); 0000-0001-6827-1024 (P. d. A. Silva)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

assessments. This integration expands the analytical capabilities of cybersecurity operations and enables higher levels of adaptive automation [4, 5]

However, the incorporation of AI into highly automated cybersecurity environments also introduces important ethical challenges . Automated decisions may directly affect users, digital services, and organizational assets, often under conditions of uncertainty, opacity, and time pressure. Risks associated with algorithmic bias, limited explainability, excessive automation, automation bias, and difficulties in assigning accountability become particularly critical when AI-generated outputs influence or trigger operational actions [6, 7, 8, 9]

Although the cybersecurity literature provides extensive operational discussions of SOAR architectures, playbooks, and response automation, existing representations are predominantly architectural in nature. They typically describe workflows, tools, and operational sequences, but often do not explicitly formalize the semantic relationships among technical artifacts, human actors, governance mechanisms, and ethical consequences. Consequently, important concepts such as decision traceability, responsibility allocation, governance oversight, and harmful impacts are often left implicit or only informally described. Besides, AI-enabled SOAR environments constitute socio-technical systems in which technical components, organizational actors, institutional constraints, and ethical properties interact dynamically. Capturing these interactions requires conceptual precision to represent the system's structure, events, dependencies, roles, and accountability relations. In this context, this paper addresses the ethical risks of combining AI with SOAR *and how they could be formally represented.*

This work is guided by the Design Science Research (DSR) methodology [10, 11]. In DSR, a research project cycle typically involves three major activities: *(1) problem investigation*, *(2) treatment design*, and *(3) artifact validation and evaluation.* In this work, the *problem investigation phase* was documented through an Ontology Requirements Specification Document (ORSD) [12, 13], which defines the competency questions and non-functional requirements that the proposed reference model must satisfy. *The treatment design phase* was conducted using OntoUML modeling principles and ontological foundations [14, 15]. The *artifact validation and evaluation* are preliminary and consist of a critical analysis of the ontology model based on cybersecurity and trustworthy AI governance principles [1, 9, 16], as well as ontology engineering and OntoUML best practices [17, 14].

The remainder of this article is organized as follows. **Section 2** presents a theoretical foundation of SOAR, and the research methodology based on Design Science is further explained in **Section 3**. The Ontology Requirements Specification Document (ORSD) is described in **Section 4**. **Section 5** presents the proposed ontology in OntoUML, and its critical analysis follows in **Section 6**. **Section 7** qualifies this work regarding companion research. Finally, **Section 8** brings some conclusions and future research directions.

2. Theoretical Foundation

Security Orchestration, Automation, and Response (SOAR) platforms have become an important component of contemporary cybersecurity operations, particularly in Security Operations Centers (SOCs), where large volumes of heterogeneous alerts must be continuously analyzed and managed. In operational environments characterized by increasing attack complexity, growing numbers of interconnected assets, and limited human resources, SOAR systems support the integration of heterogeneous security technologies, automate repetitive operational tasks, and coordinate incident response activities through predefined workflows known as *playbooks* [1].

Technically, SOAR platforms serve as orchestration mechanisms to connect multiple tools, including Security Information and Event Management (SIEM) systems, endpoint detection and response (EDR) platforms, intrusion detection systems, threat intelligence repositories, and ticketing systems. By coordinating these components, SOAR platforms contribute to faster response times, greater consistency in operational procedures, and greater scalability of security operations [18, 5].

More recently, AI techniques have been incorporated into SOAR environments. Machine learning models support alert classification, anomaly detection, incident prioritization, decision recommenda-

tions, and automated response execution [3, 4]. This combination of orchestration, automation, and intelligent analytical mechanisms has led to what may be described as AI-enabled SOAR. In such environments, operational decisions may be generated partially or fully through computational inference processes, enabling more adaptive and scalable cybersecurity operations.

The incorporation of AI into SOAR environments, however, also introduces important ethical concerns. One of the central issues involves the opacity of machine learning models. In many situations, automated classifications, recommendations, or response decisions are not fully interpretable by analysts, auditors, or affected stakeholders, limiting transparency and impairing post hoc justification processes [6, 7]. In AI-enabled SOAR environments, this lack of explainability directly affects ethical requirements associated with accountability, contestability, and human oversight, particularly when automated outputs influence operational security actions.

Another relevant ethical concern involves algorithmic bias. Training datasets may contain incomplete, distorted, or historically biased patterns capable of influencing model behavior and generating systematically unfair or erroneous outcomes [8]. In AI-enabled SOAR environments, such distortions may affect ethical dimensions related to fairness, proportionality, reliability, and justified decision-making, especially when automated prioritization mechanisms influence how incidents, assets, or users are treated during cybersecurity operations [4].

In addition, AI-enabled SOAR systems present strong interdependencies among AI models, automated workflows, operational procedures, and human actors. An incorrect inference generated by an AI model may propagate through automated playbooks and trigger operational actions without sufficient human verification. As a consequence, ethical concerns such as responsibility attribution, governance failures, insufficient oversight, and lack of accountability cannot be understood exclusively as isolated technical failures. Instead, they emerge from socio-technical interactions involving computational models, organizational procedures, institutional constraints, and human decision-making processes [9].

These concerns are increasingly reflected in contemporary AI governance frameworks. International initiatives such as the OECD Principles on Artificial Intelligence [19], the UNESCO Recommendation on the Ethics of Artificial Intelligence [20], the Ethics Guidelines for Trustworthy AI of the European Commission [16], and the NIST AI Risk Management Framework [1] emphasize principles such as transparency, accountability, human oversight, robustness, contestability, and risk-based governance. In Brazil, similar concerns appear in the ongoing debate surrounding PL 2338/2023, which proposes a regulatory framework for artificial intelligence based on risk classification and governance requirements [21].

From a conceptual perspective, AI-enabled SOAR environments can therefore be understood as socio-technical systems in which technical components, organizational structures, governance mechanisms, operational procedures, and ethical consequences dynamically interact. Capturing these interactions requires conceptual representations capable of explicitly modeling entities, relations, events, dependencies, responsibilities, and normative constraints within a unified semantic structure. In this context, ontologies become particularly relevant because they support the formal integration of technical, organizational, governance, and ethical concepts into a shared and semantically consistent vocabulary, enabling the structured analysis of accountability relations, governance dependencies, and ethical risks in AI-enabled SOAR environments.

In the field of Ontology Engineering, OntoUML represents an appropriate approach for the semi-formal specification of reference ontologies, since it provides a graphical modeling language grounded in Unified Foundational Ontology (UFO) and based on UML class diagrams. OntoUML offers philosophically grounded stereotypes that enable the construction of conceptual models understandable not only by information systems specialists but also by domain experts unfamiliar with Semantic Web languages such as OWL [14, 15]. In this work, OntoUML is adopted to model and integrate the ethical, organizational, governance, and technical dimensions associated with AI-enabled SOAR environments. More specifically, the proposed ontological approach seeks to provide a semantically grounded conceptual representation capable of supporting the identification, interpretation, and analysis of ethical risks associated with automated cybersecurity operations.

3. Research Method

This study was guided by principles of Design Science Research (DSR), a methodological paradigm that seems adequate for addressing problems of technical, organizational, and societal relevance [10, 11]. The target artifact is a reference ontology to represent AI-enabled SOAR systems to support the analysis of ethical risks associated with AI-driven automation in cybersecurity governance. A DSR-based research methodology typically proceeds through three phases: problem investigation, artifact design, and artifact evaluation. The execution of these phases is described as follows.

3.1. Problem Investigation

In DSR, problem investigation may involve methods such as literature reviews, interviews, case studies, and document analysis. In this work, however, the investigation phase was primarily conducted through the elaboration of the Ontology Requirements Specification Document (ORSD), which provided a structured characterization of the conceptual challenges associated with AI-enabled SOAR environments.

In ontology engineering, the ORSD establishes the conceptual contract that guides ontology development [12]. In particular, the competency questions defined in the ORSD were used to identify the entities, relations, governance structures, ethical concerns, and operational dependencies that should be represented by the ontology. Therefore, the ORSD simultaneously supported the definition of the ontology scope and the clarification of the conceptual problem addressed in this research.

3.2. Artifact Design

The artifact design followed established ontology engineering practices [12, 13]. Since the ORSD already supported the investigation and delimitation of the conceptual problem, the design phase started with conceptual modeling activities.

Based on the requirements identified in the ORSD, the ontology was developed as a semi-formal conceptual model expressed in OntoUML. Although ontology engineering processes frequently include subsequent operationalization stages using formal languages such as OWL, this research focuses on the conceptual specification stage, providing a reference ontology intended to support the representation and discussion of AI-enabled SOAR environments and their associated governance and ethical aspects.

3.3. Artifact Evaluation

The evaluation conducted in this study consisted of a theoretical and conceptual critical analysis of the proposed ontology. Rather than focusing on empirical validation with end users or deployment in operational environments, the evaluation emphasized the analysis of conceptual adequacy, semantic coherence, and prospective utility of the model.

The evaluation process comprised three complementary steps. First, a syntactic and semantic verification of the ontology was conducted to assess the internal consistency of the represented entities, relations, events, dependencies, and constraints. Second, the ontology was analyzed with respect to its alignment with the competency questions and conceptual requirements specified in the ORSD. Finally, the evaluation examined the adequacy of the ontology to represent relevant concepts associated with AI-enabled SOAR environments, including governance arrangements, automated decision-making processes, ethical risk categories, accountability relations, organizational roles, AI components, and control mechanisms.

4. Ontology Requirements Specification Document (ORSD)

The development of the proposed ontology was guided by an Ontology Requirements Specification Document (ORSD), which establishes the semantic boundaries of the domain and provides traceability between domain requirements and subsequent conceptual modeling decisions [12].

Table 1
Ontology Requirement Specification Document

Purpose
To build an ontology to represent the domain of AI-enabled SOAR systems, covering ethical risks associated with AI-driven automation in cybersecurity operations. The ontology should formalize the relationships among technical components, automated decision-making processes, organizational actors, governance mechanisms, and potential harmful impacts.
Scope
The ontology must cover four conceptual views: (1) *technical infrastructure*, e.g., SOAR platforms, artificial intelligence models, and training datasets; (2) *operational processes*, e.g., alerts, inference events, automation events, playbooks, and security actions; (3) *governance and supervision*, encompassing accountability structures, audit mechanisms, and compliance-related artifacts; and (4) *ethical risks*, e.g., biases, harmful impacts, and affected stakeholders.
Implementation Language
OntoUML will be used to specify the reference model. The state-of-the-art tools enable semi-automated model serialization to JSON, Turtle, and OWL.
Intended End-Users
1. Security analysts analyzing decision-making processes and operational dependencies. 2. Auditors interested in supporting traceability analysis, compliance verification, and post-incident investigation. 3. Governance and risk professionals willing to examine accountability structures and ethical implications of automated cybersecurity operations. 4. Engineers and system designers may employ the ontology as a conceptual basis for the development of safer and more governable AI-enabled security systems. 5. Scholars interested in socio-technical risks and governance challenges in AI-enabled cybersecurity environments.
Intended Uses
1. To identify and analyze ethical risks associated with AI-driven automation, e.g., opacity, bias, harmful impacts, and excessive automation. 2. Support governance-oriented reasoning by making explicit accountability structures, human oversight arrangements, and organizational control mechanisms. 3. Enable auditing and investigative activities by enabling the conceptual reconstruction of decision paths and operational consequences.
Ontology Requirements
Non-Functional Requirements
- *Findability*: The OntoUML model files (.vpp) are publicly available in an anonymized repository, allowing discovery, retrieval, and reuse through persistent online access: https://anonymous.4open.science/r/OntoUML-CF83/ - *Accessibility*: The ontology was developed using the free and permanent version of Visual Paradigm Community Edition. The model can also be accessed through other open tools supporting Visual Paradigm files and OntoUML representations. - *Interoperability*: The ontology model is represented in Visual Paradigm (.vpp) format and supports export to JSON and gUFO representations, enabling integration with Semantic Web technologies and ontology engineering tools. - *Reusability*: The ontology artifacts are distributed as open-source resources and follow OntoUML conceptual modeling principles and patterns, enabling reuse, extension, adaptation, and incorporation into future ontology engineering efforts without requiring substantial restructuring.

Functional Requirements: Competency Questions [1]

CQG1. Decision and Artificial Intelligence
CQ1. Which AI model produced a decision regarding a given alert?
CQ2. Which training data were used to train the model involved in that decision?
CQ3. Which operational action was executed as a consequence of that decision?
CQG2. Operational Processes
CQ4. Which playbook generated a specific security action?
CQ5. Which alerts triggered a given automation process?
CQG3. Governance and Accountability
CQ6. Who is responsible for a particular AI model or playbook?
CQ7. Which actions were audited and by whom?
CQG4. Risks and Impacts
CQ8. Which ethical or operational risks are associated with a given decision?
CQ9. What impacts were produced for affected users or organizational assets?

Pre-Glossary of Terms

- AIModel	- SOARPlatform	- TrainingDataset	- Alert
- Playbook	- InferenceEvent	- AutomationEvent	- SecurityAction
- Risk	- Impact	- Analyst	- ModelOwner
- PlaybookOwner	- Audit	- ComplianceReport	- GovernanceFramework

5. Ontology Model in OntoUML

OntoUML is an ontology modeling language grounded in the Unified Foundational Ontology (UFO), designed to represent domain structures and relations through explicit ontological distinctions rather than solely through syntactic modeling conventions [14, 15]. OntoUML supports the representation of the structural and dynamic dimensions of socio-technical systems while explicitly capturing their internal semantic dependencies, making it suitable for the conceptualization of AI-enabled SOAR environments. These environments involve not only technical components and operational workflows but also human actors, governance structures, accountability relations, and ethical consequences. Conventional architectural or process-oriented representations often describe such elements at a functional level but do not formally capture their semantic interdependencies. In contrast, OntoUML enables explicit representation of how technical artifacts, organizational roles, operational events, and ethical properties interact within the domain.

For this reason, OntoUML was adopted in this study as the semi-formal modeling language for the proposed ontology. Its use supports semantic precision, traceability of conceptual assumptions, and greater analytical clarity regarding automated decision-making processes and governance relationships. Figure 1 presents the ontology of AI-enabled SOAR systems.

The ontology organizes the domain into four interrelated dimensions: *technical infrastructure, operational processes, governance and supervision,* and *ethical consequences and risks.* These dimensions are not conceived as isolated layers but as interconnected analytical perspectives that jointly represent the socio-technical structure of AI-enabled cybersecurity operations. The model explicitly incorporates both structural and dynamic aspects of the domain. Structural aspects concern entities, roles, organizational arrangements, and governance artifacts, whereas dynamic aspects concern operational events, causal dependencies, and decision propagation across automated workflows.

[1] The competency questions were derived from the literature on AI governance, cybersecurity automation, and risk management, particularly from NIST [1], OECD [19], and UNESCO governance principles [20]. These competency questions guided the selection of concepts and relations represented in the conceptual model and were later used as analytical criteria during the evaluation stage.

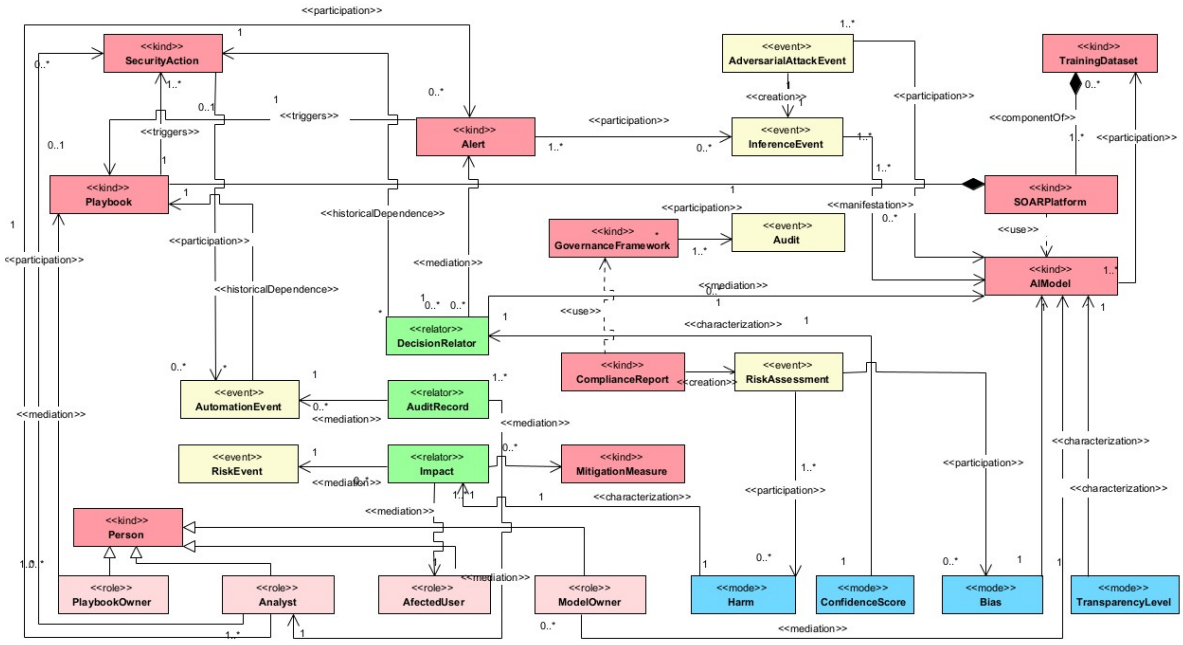


Figure 1: Proposed ontology for ethical risk analysis in AI-enabled SOAR systems

At the center of the technical dimension, SOARPlatform serves as the orchestrating entity, integrating heterogeneous cybersecurity tools and coordinating automated workflows. It constitutes the operational environment within which alerts are received, *playbooks* are activated, decisions are produced, and response actions are executed. The platform incorporates one or more AIModel instances, which support or automate decision-making activities. These models are trained using the TrainingDataset, making explicit the dependencies among data quality, model behavior, and operational outcomes. This relation is conceptually important because it allows the ontology to capture the role of training data in shaping inference behavior and, consequently, downstream operational consequences.

Operational activity begins with the occurrence of an Alert, representing a security-relevant event that triggers system activity. Alerts trigger predefined Playbook structures that define response procedures and determine how the system should react under specific operational conditions. The execution of these workflows results in SecurityAction, which represents concrete interventions in the operational environment. This sequence establishes a causal path from detection to response and provides the basis for reasoning about operational dependencies, uncertainty propagation, and potential failure points. The model's dynamic dimension is explicitly represented through InferenceEvent and AutomationEvent. InferenceEvent captures the analytical phase in which an AI model processes input data, such as alerts, and produces a decision, classification, or recommendation. AutomationEvent, in turn, represents the execution phase in which the resulting decision is translated into operational action. Explicit modeling of these events is particularly important because it enables representation of temporal ordering, causal dependencies, and propagation mechanisms. In AI-enabled SOAR environments, these dynamic relations are central to understanding how inferential uncertainty, model limitations, or data-related distortions may propagate through automated operational chains.

A central element of the ontology is the DecisionRelator. This construct provides semantic precision and traceability by mediating the relationship among AIModel, Alert, and SecurityAction. Rather than treating decision-making as an implicit internal process, the model reifies the decision as a contextualized relational entity. This ontological choice is analytically significant because it enables the explicit reconstruction of decision paths, the identification of causal dependencies, and the representation of the contextual conditions under which operational actions were produced. From a governance perspective, the DecisionRelator supports explainability, auditing, and post-incident analysis. In highly automated

environments, where decisions may otherwise remain opaque, explicit representation of the decision itself becomes essential for accountability-oriented analysis.

Human participation is explicitly represented through the class Person, which may instantiate roles such as Analyst, ModelOwner, and PlaybookOwner. These roles capture differentiated responsibilities within the socio-technical system. An Analyst represents operational oversight and human intervention in security processes; a ModelOwner is responsible for the development, maintenance, and governance of AI models; and a PlaybookOwner is responsible for workflow definition, operational logic, and response procedures. The explicit representation of these roles is particularly relevant because it makes visible the allocation of responsibility across the operational lifecycle, thereby supporting analysis of accountability structures, governance arrangements, and human oversight mechanisms in AI-enabled automation contexts.

The governance dimension is further represented by Audit, ComplianceReport, and GovernanceFramework. Audit supports systematic evaluation of operational behavior and decision pathways. ComplianceReport documents the outcomes of verification and risk assessment activities. GovernanceFramework defines the normative context within which the system operates, including organizational policies, standards, and regulatory requirements. Together, these elements provide the institutional context necessary for representing how automated cybersecurity operations are monitored, evaluated, and aligned with organizational and legal expectations.

Ethical concerns are incorporated through dependent properties (modes), including Bias, TransparencyLevel, and Harm. Rather than being represented as isolated attributes, these elements are modeled as properties that inhere in other entities. This ontological treatment enables ethical risks to be understood as emergent properties of socio-technical interactions rather than merely as abstract external concerns. For example, Bias is modeled as a property of AIModel, reflecting the possibility that distortions embedded in the data or model design may influence inferential outcomes. TransparencyLevel expresses explainability-related characteristics of automated decision processes, whereas Harm refers to the impacts of system actions on affected users, services, or organizational assets. This modeling choice allows ethical risks to be analyzed as phenomena arising from the interaction among technical components, operational conditions, organizational decisions, and affected stakeholders.

Within the proposed ontology, ethics-by-design is incorporated as an intrinsic modeling principle rather than treated as an external normative layer. This is achieved through explicit representation of decision-making processes, traceability of actions, allocation of responsibilities, modeling of governance artifacts, and representation of harmful impacts. Consequently, ethical considerations are embedded directly into the domain's conceptual structure. This modeling strategy strengthens the analytical usefulness of the ontology by enabling examination not only of how automated cybersecurity systems operate, but also of how ethical risks emerge, propagate, and may be governed. A complementary description of the modeled entities, relations, ontological stereotypes, and semantic constraints is provided in Appendix A, which contains the corresponding descriptive table and element definitions.

6. Critical Analysis of the Model and Response to the Research Questions

The evaluation of the proposed ontology was conducted through a critical conceptual analysis undertaken as an independent methodological procedure rather than as an interpretive discussion of results. In conceptual modeling research, critical analysis plays an important evaluative role because it enables systematic examination of whether an artifact adequately represents the relevant entities, relations, assumptions, and explanatory mechanisms of the target domain [10, 11]. In this study, the analysis was guided by two complementary analytical bases. First, the ontological foundations of OntoUML and the Unified Foundational Ontology (UFO) were adopted as conceptual criteria for assessing semantic coherence, ontological completeness, and the adequacy of the relations represented in the model [14, 15]. This perspective enabled examination of whether the conceptual distinctions adopted in the ontology, such as kinds, roles, relators, events, and dependent properties, were appropriate for

representing the semantics of AI-enabled SOAR environments. Second, the analysis was informed by governance-oriented conceptual principles derived from the literature on trustworthy and responsible artificial intelligence. Explainability, accountability, traceability, human oversight, and risk governance were considered as relevant analytical criteria [16, 1]. Together, these two analytical lenses provided the conceptual basis for evaluating the adequacy of the proposed ontology and its ability to address the research question.

The central research question of this study was *which requirements an ontology for the AI-enabled SOAR domain should satisfy to represent and support the analysis of ethical risks associated with AI-driven automation.* The critical analysis indicates that such an ontology must explicitly represent the domain's structural components, including technical artifacts, human actors, organizational roles, and governance mechanisms. It must also capture the dynamic aspects of automated cybersecurity operations, including events, causal dependencies, operational sequencing, and decision propagation. In addition, it must explicitly represent accountability structures and traceability relations so that decision pathways, responsibilities, and governance mechanisms become analytically visible. Finally, ethical properties and harmful impacts must be incorporated as integral elements of the conceptual representation rather than treated as external normative considerations. The proposed ontology addresses these requirements by explicitly integrating technical infrastructure, operational dynamics, governance artifacts, and ethical consequences within a single conceptual framework.

Regarding the technical risks arising from the integration of artificial intelligence and SOAR, the analysis indicates that many risks do not emerge from isolated components but rather from interactions among heterogeneous elements of the socio-technical system. For example, adversarial manipulation, incomplete training data, or low-quality operational inputs may affect the behavior of an AIModel, leading to incorrect inferences during an InferenceEvent. Once incorporated into automated playbooks, these erroneous outputs may trigger inappropriate SecurityAction. In this way, an initial inferential distortion may propagate through interconnected operational chains, with broader consequences. The ontology makes this propagation analytically explicit by representing alerts, inference events, automation events, playbooks, and resulting actions as causally related components of a structured operational sequence. This constitutes an important contribution because it allows technical failures to be examined not only as isolated computational errors but also as systemic mechanisms of propagation within automated cybersecurity environments.

The critical analysis also highlights the importance of human supervision as a governance mechanism. By explicitly representing the roles of Analyst, ModelOwner, and PlaybookOwner, the ontology makes the distribution of responsibilities across the operational lifecycle visible. This representation is conceptually relevant because governance in AI-enabled cybersecurity systems depends not only on the quality of automated decisions but also on the institutional allocation of authority, responsibility, and intervention capacity. This finding is consistent with the principle of human oversight emphasized in contemporary AI governance literature, according to which automated systems should remain subject to meaningful human intervention, particularly when decisions may generate significant organizational or societal consequences [16]. From this perspective, the model supports the analysis of human-in-the-loop arrangements as conceptual mechanisms for reducing excessive automation, improving organizational control, and strengthening governance accountability.

Regarding ethical challenges, the model indicates that negative consequences may arise not only from technical malfunctions but also from the intrinsic properties of AI systems and the socio-technical conditions under which they operate. The inclusion of dependent properties such as Bias, TransparencyLevel, and Harm enables explicit representation of how ethical risks emerge from interactions among model behavior, operational context, organizational decisions, and affected stakeholders. This conceptual treatment is particularly important because it avoids reducing ethical concerns to isolated attributes of technical artifacts. Instead, ethical risks are represented as emergent properties of socio-technical relations. This perspective is consistent with contemporary responsible AI approaches, which emphasize that risks associated with algorithmic opacity, unfairness, and unintended harmful impacts should be understood as socio-technical phenomena rather than merely as technical defects [1].

Another important result of the critical analysis concerns traceability and explainability. The explicit

representation of DecisionRelator, together with Audit and ComplianceReport, provides conceptual mechanisms for reconstructing the sequence of events leading to specific automated actions. From the perspective of governance and accountability, this capability is particularly relevant because it supports post-incident investigation, retrospective examination of decision pathways, and regulatory verification. In highly automated operational environments, where decisions may otherwise remain opaque, the explicit conceptual representation of decision-making itself constitutes a significant analytical advantage.

Despite these contributions, the critical analysis also identifies important limitations. The proposed artifact remains a semi-formal conceptual ontology and has not yet been empirically validated in operational environments. As emphasized in Design Science Research, conceptual adequacy alone does not guarantee practical applicability; empirical evaluation in real organizational contexts is required [10]. In addition, cybersecurity ecosystems evolve rapidly. New attack patterns, new classes of AI models, emerging governance requirements, and domain-specific regulatory constraints may require future conceptual extensions. Moreover, the present study focuses on conceptual modeling rather than computational implementation. Consequently, questions related to formal reasoning performance, interoperability with operational security infrastructures, and integration with existing cybersecurity knowledge bases remain outside the scope of the present work.

Overall, the critical analysis indicates that the proposed ontology provides a coherent conceptual foundation for addressing the research question. By combining ontological rigor with governance-oriented analytical criteria, the model enables a structured examination of how artificial intelligence, operational automation, human responsibility, and ethical consequences interact in AI-enabled SOAR environments. In this sense, the ontology contributes not only as a representational artifact but also as an analytical instrument for ethical risk examination, governance-oriented reasoning, and future empirical investigation in increasingly automated cybersecurity systems.

7. Related Work

Security Orchestration, Automation, and Response (SOAR) has been widely discussed in the cybersecurity literature as an architectural and operational approach for integrating heterogeneous security tools, automating repetitive tasks, and coordinating incident response activities within Security Operations Centers. Existing studies generally emphasize operational architecture, workflow orchestration, alert prioritization, and performance-related outcomes such as reduced response time, increased analyst productivity, and improved scalability of security operations [18] [1].

These contributions provide important practical and organizational perspectives. However, most existing representations remain predominantly architectural, process-oriented, or implementation-driven. They typically describe tools, connectors, event pipelines, playbooks, and operational sequences, but often do not explicitly formalize the semantic distinctions among entities, roles, events, dependencies, and governance mechanisms. Therefore, important concepts such as automated decision-making, accountability structures, human oversight, ethical impacts, and harmful consequences are often implicit or represented only informally.

Recent studies addressing artificial intelligence in cybersecurity have also explored machine learning techniques for alert triage, anomaly detection, predictive analytics, and automated incident response. However, these studies generally focus on algorithmic performance, detection accuracy, operational efficiency, or implementation feasibility rather than on explicit conceptual formalization of socio-technical relationships and ethical implications [6, 1]. More specifically, the literature rarely provides conceptual models that formally represent how AI models, automated decisions, organizational actors, governance structures, and ethical risks interact within AI-enabled SOAR environments.

From a conceptual modeling perspective, ontology-driven approaches have increasingly been adopted to improve semantic precision in complex socio-technical domains. OntoUML, grounded in the Unified Foundational Ontology (UFO), has been applied in areas such as governance, enterprise architecture, healthcare, legal compliance, and organizational modeling to explicitly distinguish kinds, roles, relators, events, and dependent properties [14, 15].

These studies demonstrate that ontological modeling can reduce conceptual ambiguity, improve semantic consistency, and strengthen the conceptual foundations required for communication, analysis, and future computational implementation. The main contribution of the present study lies precisely in advancing this perspective. Rather than proposing an additional operational or architectural representation of SOAR, this paper introduces a semi-formal conceptual ontology of AI-enabled SOAR systems expressed in OntoUML. By grounding the model in explicit ontological categories and semantically well-defined relations, the proposed approach reduces the interpretive subjectivity often found in narrative, process-oriented, or purely diagrammatic representations, while providing a more rigorous conceptual basis for analyzing governance structures, automated decision-making processes, and ethical risks.

8. Conclusions

This paper proposed a conceptual ontology for AI-enabled SOAR systems with a particular focus on ethical risk analysis, governance, and accountability in automated cybersecurity environments. Grounded in an Ontology Requirements Specification Document (ORSD) and modeled in OntoUML, the proposed artifact provides a semi-formal conceptual representation of the relationships among technical infrastructure, operational processes, human roles, governance mechanisms, and ethical consequences.

The study contributes to the literature in two complementary ways. First, it provides a structured conceptual representation of AI-enabled SOAR as a socio-technical system in which automated decisions emerge from the interaction among data, models, operational workflows, human actors, and institutional constraints. This perspective makes explicit the interconnected nature of technical, organizational, and governance dimensions that shape cybersecurity operations. Second, by adopting an ontologically grounded modeling language, the study introduces a level of semantic precision that is often absent from existing architectural and process-oriented descriptions of SOAR. This enables the explicit representation of decision traceability, accountability relations, governance structures, and ethical consequences associated with automated security actions, thereby offering a more rigorous conceptual basis for the analysis and governance of AI-driven cybersecurity automation.

The critical analysis conducted in this research indicates that the ontology adequately addresses the research question by explicitly specifying the conceptual requirements for representing ethical risks in AI-enabled cybersecurity automation. In particular, the model supports examination of how technical failures may propagate through automated operational chains, how human oversight can be incorporated into governance structures, and how ethical risks such as bias, opacity, and harmful impacts may emerge from socio-technical interactions.

The study also presents limitations. The proposed artifact remains a semi-formal conceptual model and has not yet been empirically validated in operational environments. In addition, the evolving nature of cybersecurity ecosystems suggests that future extensions may be required to represent emerging attack patterns, evolving artificial intelligence techniques, and domain-specific regulatory developments.

Future research may advance this work in three complementary directions. First, the ontology may be formalized in languages such as OWL. Second, empirical validation may be conducted with cybersecurity practitioners, auditors, and governance specialists. Third, the model may be applied in real or simulated SOAR environments to support auditing, ethical risk assessment, governance-oriented reasoning, and decision analysis.

Overall, the proposed ontology provides a semantically grounded conceptual foundation for understanding how artificial intelligence, operational automation, human responsibility, and ethical consequences interact in contemporary cybersecurity environments. In this sense, it contributes to the development of more transparent, accountable, and responsible approaches to AI-enabled security orchestration and response.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-4 for translation review and grammar and spelling checking. After using this tool, the authors reviewed and edited the content as needed and assume full responsibility for the content of the publication.

References

[1] National Institute of Standards and Technology, Artificial intelligence risk management framework (ai rmf 1.0), 2023. URL: https://doi.org/10.6028/NIST.AI.100-1. doi:10.6028/NIST.AI.100-1.

[2] D. Kremer, et al., Intelligent system for assisting security experts in generating playbooks for automated incident response, in: Proceedings of the International Conference on Security and Cryptography (SECRYPT), SciTePress, Rome, Italy, 2023. URL: https://arxiv.org/abs/2311.03825, accessed: 2026-05-22.

[3] Vectra AI, Incident response automation, 2026. URL: https://www.vectra.ai/topics/incident-response-automation, accessed: 2026-05-22.

[4] Jose Enrique, Ai-powered soc guide: Artificial intelligence in modern security operations centers, 2026. URL: https://ciberseguridad.joseenrique.es/en/article-ai-soc.html, accessed: 2026-05-22.

[5] CrowdStrike, Security orchestration, automation and response (soar), 2026. URL: https://www.crowdstrike.com/en-us/cybersecurity-101/next-gen-siem/security-orchestration-automation-and-response-soar/, accessed: 2026-05-22.

[6] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (2019) 206–215. doi:10.1038/s42256-019-0048-x.

[7] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608 (2017). URL: https://arxiv.org/abs/1702.08608. arXiv:1702.08608.

[8] S. Barocas, M. Hardt, A. Narayanan, Fairness and Machine Learning: Limitations and Opportunities, MIT Press, Cambridge, MA, 2019. URL: https://fairmlbook.org.

[9] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schäfer, P. Valcke, E. Vayena, Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations, Minds and Machines 28 (2018) 689–707. doi:10.1007/s11023-018-9482-5.

[10] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS Quarterly 28 (2004) 75–105. doi:10.2307/25148625.

[11] S. Gregor, A. R. Hevner, Positioning and presenting design science research for maximum impact, MIS Quarterly 37 (2013) 337–355.

[12] M. C. Suárez-Figueroa, A. Gómez-Pérez, B. Villazón-Terrazas, How to write and use the ontology requirements specification document, in: R. Meersman, T. Dillon, P. Herrero (Eds.), On the Move to Meaningful Internet Systems: OTM 2009 Workshops, volume 5872 of *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2009, pp. 966–982. doi:10.1007/978-3-642-05290-3_84.

[13] M. C. Suárez-Figueroa, A. Gómez-Pérez, Ontology requirements specification, in: M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, A. Gangemi (Eds.), Ontology Engineering in a Networked World, volume 11 of *Studies on the Semantic Web*, Springer, Berlin, Heidelberg, 2012, pp. 93–106. doi:10.1007/978-3-642-24794-1_5.

[14] G. Guizzardi, Ontological Foundations for Structural Conceptual Models, Ph.D. thesis, University of Twente, Enschede, 2005. URL: https://research.utwente.nl/en/publications/ontological-foundations-for-structural-conceptual-models.

[15] G. Guizzardi, J. P. A. Almeida, D. Porello, V. A. Carvalho, T. P. Sales, OntoUML 2.0 Specification, Technical Report, Ontology and Conceptual Modeling Research Group, Vitória, Brazil, 2022. URL: https://ontouml.readthedocs.io/en/latest/.

[16] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, Technical Report, European Commission, Brussels, 2019. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[17] T. R. Gruber, Toward principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer Studies 43 (1995) 907–928. doi:10.1006/ijhc.1995.1081.
[18] D. Shackleford, Security Orchestration, Automation and Response (SOAR), White Paper, SANS Institute, Bethesda, MD, 2018. URL: https://www.sans.org/white-papers/security-orchestration-automation-response-soar/.
[19] Organisation for Economic Co-operation and Development, Oecd principles on artificial intelligence, 2019. URL: https://doi.org/10.1787/eb4e399d-en. doi:10.1787/eb4e399d-en.
[20] United Nations Educational, Scientific and Cultural Organization, Recommendation on the ethics of artificial intelligence, 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000381137.
[21] Brasil, Projeto de lei nº 2.338, de 2023: Dispõe sobre o uso da inteligência artificial no brasil, Projeto de Lei, 2023. URL: https://www.camara.leg.br/propostas-legislativas/2352175, acesso em: 5 maio 2026.

A. OntoUML Relationships Table

Domain Class	Description	Stereotype	Main Relationships
SOARPlatform	Central platform responsible for orchestrating automation and integrating security and AI tools	«kind»	hasPart → Playbook; uses → AIModel
AIModel	AI model used for analysis, classification, and automated decision-making	«kind»	trainedWith → TrainingDataset; participates → InferenceEvent; mediates (via DecisionRelator)
TrainingDataset	Dataset used for training the AI model	«kind»	provides data to → AIModel
Playbook	Automated workflow that defines incident response actions	«kind»	triggers → SecurityAction; definedBy → AutomationEvent; partOf → SOARPlatform
Alert	Security event or notification that initiates the operational process	«kind»	triggers → Playbook; analyzedBy → InferenceEvent
SecurityAction	Action executed in response to an alert	«kind»	executedBy → AutomationEvent; justifiedBy → DecisionRelator
InferenceEvent	Event in which the AI model analyzes an alert and generates an inference	«event»	executedBy → AIModel; analyzes → Alert
AutomationEvent	Event responsible for executing automated actions	«event»	executes → SecurityAction; definedBy → Playbook
AdversarialAttackEvent	Event representing attacks against the AI model	«event»	targets → AIModel; mayCause → InferenceError
RiskAssessment	Process for assessing risks related to the system	«event»	produces → ComplianceReport; identifies → Bias, Harm
Audit	Audit event for compliance verification and traceability	«event»	verifies → GovernanceFramework; generates → AuditRecord
GovernanceFramework	Governance structure that defines system rules and policies	«kind»	verifiedBy → Audit
ComplianceReport	Report generated from risk assessments and audits	«kind»	producedBy → RiskAssessment
MitigationMeasure	Measure applied to mitigate identified negative impacts	«kind»	mitigates → Impact
Person	Represents human agents within the system	«kind»	plays → Analyst, ModelOwner, PlaybookOwner, AffectedUser
Analyst	Responsible for analyzing alerts and reviewing actions	«role»	reviews → SecurityAction; investigates → Alert; participates → AuditRecord
ModelOwner	Responsible for AI model management and performance	«role»	responsibleFor → AIModel
PlaybookOwner	Responsible for playbooks and automations	«role»	responsibleFor → Playbook
AffectedUser	User impacted by automated decisions	«role»	participates → Impact
DecisionRelator	Entity that connects decision, model, alert, and action, ensuring traceability	«relator»	mediates → AIModel, Alert; justifies → SecurityAction
AuditRecord	Audit record that links actions and responsible parties	«relator»	mediates → AutomationEvent, Analyst
Impact	Relates risk events to affected users	«relator»	mediates → RiskEvent, AffectedUser
Bias	Represents bias present in the AI model	«mode»	inheresIn → AIModel; identifiedBy → RiskAssessment
TransparencyLevel	Degree of explainability of the AI model	«mode»	inheresIn → AIModel
ConfidenceScore	Level of confidence associated with a decision	«mode»	inheresIn → DecisionRelator
Harm	Represents harm caused by automated decisions	«mode»	inheresIn → Impact; identifiedBy → RiskAssessment