

Titre: Outil d'aide à la décision pour la gestion de la conformité des TI
Title:

Auteur: Marwane Adala
Author:

Date: 2023

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Adala, M. (2023). Outil d'aide à la décision pour la gestion de la conformité des TI
Citation: [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
<https://publications.polymtl.ca/53461/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/53461/>
PolyPublie URL:

Directeurs de recherche: Nora Boulahia Cuppens, & Frédéric Cuppens
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Outil d'aide à la décision pour la gestion de la conformité des TI

MARWANE ADALA

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie informatique

Juin 2023

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Outil d'aide à la décision pour la gestion de la conformité des TI

présenté par **Marwane ADALA**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Martine BELLAÏCHE, présidente

Nora BOULAHIA CUPPENS, membre et directrice de recherche

Frédéric CUPPENS, membre et codirecteur de recherche

Benoît ROBERT, membre

DÉDICACE

*À tous mes amis du labos,
vous me manquerez...*

REMERCIEMENTS

J'ai le plaisir d'exprimer ma gratitude à tous ceux qui m'ont soutenu, guidé et aidé d'une manière ou d'une autre au cours de ce projet, que ce soit directement ou indirectement. Plus particulièrement, je tiens à exprimer ma gratitude à Nora-Boulahia Cuppens, PhD, Frédéric Cuppens, PhD et Samra Bouakkaz, PhD pour m'avoir supervisé et pour m'avoir accompagné, encouragé et conseillé durant ces mois. Je tiens à remercier le partenaire industriel ainsi que tous les autres membres du laboratoire d'accueil pour leurs conseils, leurs encouragements et pour la sympathie qu'ils m'ont témoignée.

Merci à vous tous !

RÉSUMÉ

La Gouvernance, les Risques et la Conformité (GRC) en Technologie de l'Information (TI) sont devenus une priorité pour les organisations, particulièrement la gestion de la conformité, qui présente d'importants défis comme l'automatisation de sa vérification. Dans cette optique, notre mémoire de maîtrise, réalisé en collaboration avec un partenaire industriel, propose un outil automatisé d'aide à la décision pour la gestion de la conformité des TI dédié aux Petites et Moyennes Entreprises (PME). Notre travail s'est articulé autour de plusieurs axes : la définition des métriques et indicateurs de conformité, la création d'un outil intégrant les règles et contrôles de conformité dans une ontologie sémantique, l'élaboration d'un pipeline d'apprentissage automatique pour anticiper la conformité des journaux et analyses de sécurité, et enfin, la production d'un prototype fonctionnel fournissant des recommandations pour résoudre les problèmes de conformité. Pour la conception, le développement et l'évaluation de notre solution, nous avons adopté les approches Action Design Research (ADR) et Design Science Research (DSR). Une revue exhaustive de la littérature a permis de collecter des données essentielles et de construire une base de données des métriques de conformité. Inspiré des outils open source, notre outil se compose de trois modules : Module de Collecte de Données, Module de Vérification de Conformité et Module de Tableaux de Bord. L'ontologie développée dans ce projet a été validée par des experts du domaine grâce à un ensemble de questions de compétence basées sur les cas d'utilisation. Notre solution répond aux défis des PME en matière de gestion de la conformité des TI, contribue aux recherches sur la modélisation des connaissances, l'intelligence artificielle et la cybersécurité, et ouvre la voie à de nouvelles innovations dans ces derniers domaines, bénéficiant à terme aux organisations de tous secteurs et tailles.

ABSTRACT

Governance, Risk and Compliance (GRC) in Information Technology (IT) has become a priority for organizations, especially compliance management, which presents important challenges such as automating its verification. In this perspective, our master's thesis, realized in collaboration with an industrial partner, proposes an automated decision support tool for IT compliance management dedicated to Small and Medium Enterprises (SME). Our work was articulated around several axes: the definition of compliance metrics and indicators, the creation of a tool integrating compliance rules and controls in a semantic ontology, the development of a machine learning pipeline to anticipate compliance in security logs and analyses, and finally, the production of a functional prototype providing recommendations to solve compliance problems. For the design, development and evaluation of our solution, we adopted the Action Design Research (ADR) and Design Science Research (DSR) approaches. A comprehensive literature review was used to collect essential data and build a database of compliance metrics. Inspired by open source tools, our solution consists of three modules: Data Collection Module, Compliance Verification Module and Dashboard Module. The ontology developed in this project has been validated by domain experts through a set of competency questions based on use cases. Our solution addresses the IT compliance management challenges of small and medium-sized businesses, contributes to research on knowledge modeling, artificial intelligence and cybersecurity, and paves the way for new innovations in these latter areas, ultimately benefiting organizations of all sectors and sizes.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	xi
LISTE DES FIGURES	xii
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
LISTE DES ANNEXES	xvii
CHAPITRE 1 INTRODUCTION	1
1.1 Définitions des notions de base	1
1.1.1 Conformité et Conformité Informatique	1
1.1.2 Catalogues de contrôles et audit de sécurité	2
1.1.3 Ontologies	2
1.2 Problématique	3
1.3 Objectifs de la Recherche	4
1.3.1 Objectif Principal de la Recherche	4
1.3.2 Sous-objectifs de la Recherche	4
1.3.3 Finalités de l'étude et Intérêt pour le domaine	4
1.4 Définition du champ d'application	5
1.4.1 Sources de données et catégories	6
1.5 Considérations éthiques	7
1.6 Plan du mémoire	7
CHAPITRE 2 REVUE DE LA LITTÉRATURE	9
2.1 Revue de la littérature sur les réglementations et les normes	9
2.1.1 Automatisation de la conformité	9

2.1.2	Métriques, paramètres et indicateurs clés de la conformité	13
2.1.3	Conformité informatique et Politiques de Sécurité de l'Information (PSI)	15
2.1.4	Facteurs affectant la gestion de la conformité	16
2.1.5	Aperçu du marché des solutions de conformité	16
2.2	Revue de la littérature sur la modélisation des connaissances	18
2.2.1	Classifications de la connaissance	18
2.2.2	Web Ontology Language (OWL)	19
2.2.3	Utilisation et pertinence des ontologies	20
2.3	Illusions des normes de conformité informatique	23
2.3.1	Défis et perspectives dans la conformité informatique	23
2.3.2	Regroupement et analyse des journaux	24
2.4	Discussion	25
 CHAPITRE 3 DÉVELOPPEMENT DE LA SOLUTION DE GESTION DE CONFORMITÉ DES TI		
3.1	Méthodologie de recherche	27
3.1.1	Étape 1 : Formulation du problème	28
3.1.2	Étape 2 : Construction, Mise en œuvre et Évaluation (CMÉ)	29
3.1.3	Étape 3 : Réflexion et apprentissage	31
3.1.4	Étape 4 : Formalisation de l'apprentissage	32
3.2	Architecture de la solution	33
3.2.1	Approche méthodologique pour l'architecture de la solution	33
3.2.2	Spécificités de l'architecture du projet	33
3.3	Module de Collecte de Données (MCD)	35
3.3.1	Élaboration du Module de Collecte de Données	35
3.3.2	Contrôles de sécurité sélectionnés	35
3.3.3	Données fournies par le partenaire industriel	36
3.3.4	Génération de journaux avec Logger	36
3.3.5	Adaptation des règles Wazuh pour la conformité	37
3.3.6	Utilisation d'ontologies open-source pour la conformité	38
 CHAPITRE 4 MODULE DE VÉRIFICATION DE CONFORMITÉ (MVC)		
4.1	La Base de Connaissances Ontologiques (BCO)	42
4.1.1	Objectifs et portée de l'ontologie	43
4.1.2	Acquisition et structuration des connaissances	43
4.1.3	Formalisation et optimisation de l'ontologie	44
4.1.4	Structure détaillée de l'ontologie	44

4.2	L'Analyseur des Journaux (AJ)	46
4.2.1	Modèle d'Analyse de Conformité (MAC)	49
4.2.2	Jeu de Règles de Conformité (JRC)	51
4.2.3	Mécanisme de Regroupement pour la Conformité (MRC)	54
4.2.4	Générateur de Rapports et de Corrélations de Conformité (GRCC)	55
4.3	Base de Données des Métriques de Conformité (BDMC)	57
4.3.1	Application du paradigme Goal-Question-Metric (GQM)	57
4.3.2	Terminologie et description de BDMC	58
4.3.3	Taxonomie des métriques identifiées	58
4.3.4	Présentation de la méthodologie concernant l'utilisation des métriques	60
4.4	Optimisation et évaluation de l'Analyseur de Journaux	61
CHAPITRE 5 OPTIMISATION DE L'INTERPRÉTATION ET DE LA VISUALISATION DE LA POSTURE DE CONFORMITÉ		65
5.1	Intérêt de l'AIOS	65
5.2	Pipeline d'apprentissage automatique utilisant ATOM	66
5.3	La méthodologie OnML	73
5.4	Amélioration de l'interprétabilité du pipeline ATOM	75
5.5	Bonification de l'ontologie	76
5.6	Module de Tableau de Bord (MTB)	78
5.6.1	Interface client et soumission des journaux informatiques	78
5.6.2	Rapport d'évaluation de la conformité	79
5.6.3	Visualisations et recommandations	79
CHAPITRE 6 VALIDATION DE L'ONTOLOGIE BONIFIÉE		85
6.1	Méthodologie de validation	85
6.2	Procédure de validation	86
6.3	Aperçu des questions de compétences	87
CHAPITRE 7 CONCLUSION ET PERSPECTIVES		92
7.1	Synthèse des travaux	92
7.2	Menaces pour la validité	93
7.3	Limitations de la solution proposée	94
7.3.1	Limites inhérentes	94
7.3.2	Fonctionnalités incluses dans la portée	95
7.3.3	Fonctionnalités hors du champ d'application	95
7.4	Améliorations et Recherches futures	95

RÉFÉRENCES	98
ANNEXES	117

LISTE DES TABLEAUX

Tableau 3.1	Les activités de l'Étape 1	29
Tableau 3.2	Les activités de l'Étape 2	31
Tableau 3.3	Les activités de l'Étape 3.	32
Tableau 3.4	Les activités de l'Étape 4.	32
Tableau 4.1	Catégories de haut niveau des métriques identifiées	59
Tableau 4.2	Métriques de conformité liées à l'analyse de journaux	61
Tableau 4.3	Hyperparamètres optimaux pour l'Analyseur de Journaux	62
Tableau 4.4	Résultats des métriques pour les phases d'entraînement, de validation et de test	64
Tableau 5.1	Évaluation de la conformité par rapport à PCI-DSS	79
Tableau 5.2	Évaluation de la conformité par rapport à NIST-SP 800-53	79
Tableau 5.3	Évaluation de la conformité par rapport à ISO27002	80
Tableau 5.4	Recommandations basées sur l'évaluation de la conformité et contrôles relatifs	84

LISTE DES FIGURES

Figure 2.1	Cycle de Hype Gartner pour les technologies juridiques et de conformité	17
Figure 3.1	Description du cycle Construction, Mise en Oeuvre et Évaluation (CMÉ) liés à ce projet	30
Figure 3.2	Architecture globale du projet	33
Figure 4.1	Composants du Module de Vérification de Conformité (MVC).	42
Figure 4.2	Aperçu du fichier Excel utilisé pour structurer les normes de sécurité de l'information	45
Figure 4.3	Pipeline de l'Analyseur des Journaux	47
Figure 4.4	Exemple d'analyse d'une ligne de journaux par le MAC	49
Figure 4.5	Le flux de travail de Jeu de Règles de Conformité (JRC)	53
Figure 4.6	Le Mécanisme de Regroupement pour la Conformité (MRC)	54
Figure 4.7	Évaluation des métriques et des hyperparamètres	63
Figure 5.1	Le Pipeline d'Automated Tool for Optimized Modelling (ATOM)	67
Figure 5.2	Résultats des modèles choisis dans le framework ATOM	68
Figure 5.3	Matrice de confusion du modèle Passive Agressive	73
Figure 5.4	Organigramme de l'approche OnML	74
Figure 5.5	Schéma du pipeline interprétable	77
Figure 5.6	Interface client pour tester l'outil	78
Figure 5.7	Conformité par rapport PCI_Requirement_1.1.1	81
Figure 5.8	Conformité par rapport NIST_SP-800-53_AC-2	82
Figure 5.9	Conformité par rapport ISO27002_9.0	83
Figure .1	Carte de la littérature avec LitMap	118
Figure .2	Sources et dossiers bibliographiques dans Zotero	119
Figure .3	Tableau représentant les chaînes de recherche utilisées pour un besoin d'information spécifique	120
Figure .4	Commande de recherche d'experts utilisée dans "Engineering Village"	121
Figure .5	Hiérarchie des classes et individus de l'ontologie.	122
Figure .6	Classes liées aux actifs, aux menaces, aux vulnérabilités et aux contrôles	125
Figure .7	Classes liées aux utilisateurs et au contrôle d'accès	126
Figure .8	Classes de contrôle de sécurité, d'architecture et de données	127
Figure .9	Classes liées à la protection de la vie privée et à la réglementation	128

LISTE DES SIGLES ET ABRÉVIATIONS

AAC	Analyse Avancée des Contrats
AA	Apprentissage Automatique
AcS	Actual Severity
AC	Anomalies de Corrélacion
ADR	Action Design Research
AECID	Automatic Event Correlation for Incident Detection
AF	Anomalies de Fréquence
AIOS	Agent de l'Intelligence artificielle Orienté Sémantique
AJC	Analyse Juridique et de Conformité
AJ	Analyseur des Journaux
AMiner	Logdata Anomaly Miner
AMO	Access Management Ontology
AMSS	Adaptive Mapping of Security Standard ontology
API	Application Programming Interface
AQ	Assurance de la Qualité
ASCII	American Standard Code for Information Interchange
ASCT	Automated Security Compliance Tool
AS	Anomalies de Séquence
ATA	Anomalies dans le Temps d'Arrivée
ATOM	Automated Tool for Optimized Modelling
AWS	Amazon Web Services
BACLO	Basic Access Control Ontology
BCO	Base de Connaissances Ontologiques
BDMC	Base de Données de Métriques de Conformité
BIE	Building, Implementing and Evaluation
BIM	Building Information Modeling
BYOD	Bring Your Own Device
CCM	Cloud Control Matrix
CCPA	California Consumer Privacy Act
CEPI	Compliance Expense Per Incident
CH	Compliance Health
CIA	Confidentiality, Integrity and Availability
CIS	Center for Internet Security

CMMI	Capability Maturity Model Integration
CMÉ	Construction, Mise en Oeuvre et Évaluation
COBIT	Control Objectives for Information and Related Technologies
COPPA	Children’s Online Privacy Protection Act
CSAT-RBAC	Context Sensitive Attribute and Task- Role Based Access Control
CSA	Cloud Security Alliance
CTP	CloudTrust
CU	Comportement de l’Utilisateur
DevOps	Development and Operations
DL	Description Logic
DSR	Design Science Research
D	Distance
EIVP	Évaluations de l’Impact sur la Vie Privée
FAIR	Findable, Accesible, Interoperable, and Reusable
FBRAM	Frame-Based Requirements Analysis Method
FedRAMP	Federal Risk and Authorization Management Program
FERPA	Family Educational Rights and Privacy
FIBO	Financial Industry Business Ontology
FISMA	Federal Information Security Management Act
FOAF	Friend-Of-A-Friend
GCVC	Gestion du Cycle de Vie des Contrats
GDPR	General Data Protection Regulation
GDR	Gouvernance Dynamique des Risques
GIR	Gestion Intégrée des Risques
GLBA	Gramm-Leach-Bliley Act
GQM	Goal-Question-Metric
GRCC	Générateur de Rapports et de Corrélations de Conformité
GRC	Gouvernance, Risque et Conformité
GRLT	Gestion des Risques Liés aux Tiers
HIPAA	Health Insurance Portability and Accountability Act
HTML	HyperText Markup Language
HTTP	Hyper Text Transfer Protocol
IA	Intelligence Artificielle
IaaS	Infrastructure-as-a-Service
IDT	Incident Discovery Time
IE	Information Extraction

IRS P1075	Internal Revenue Service Publication 1075
IRT	Incident Resolution Time
ISO/IEC	The International Organization for Standardization/ The International Electrotechnical Commission
ISO	International Organization for Standardization
ITSMO	IT Service Management Ontology
IT	Information Technology
JRC	Jeu de Règles de Conformité
KCI	Key Compliance Indicators
KRI	Key Risk Indicator
LKIF	Legal Knowledge Interchange Format
MAC	Modèle d'Analyse de Conformité
MCD	Module de Collecte de Données
MeLOn	Methodology for building Legal Ontology
ML	Machine Learning
MMT	Mitigation Mean Time
MRC	Mécanisme de Regroupement pour la Conformité
MTB	Module de Tableaux de Bord
MTID	Mean Time for Incident Discovery
MTIR	Mean Time for Incident Resolution
MVC	Module de Vérification de Conformité
NCI	Number of Compliance Incidents
NERC CIP	North American Electric Reliability Corporation Critical Infrastructure Protection
NIST SP 800-53	National Institute of Standards and Technology- Security and Privacy Controls for Information Systems and Organizations
NIST	National Institute of Standards and Technology
NLP	Nature Language Processing
OBACM	Ontology Based Access Control Model
OBAC	Ontology Based Access Control
OdTM	Ontology-driven Threat Modeling
OLLIE	Open Language Learning for Information Extraction
OnML	Ontology-based Interpretable Machine Learning
OpenBIM	Open Building Information Modeling
OntoHCCS	Ontology for a HIPAA Compliant Cloud Service
OWL	Web Ontology Language

PCI-DSS	Payment Card Industry- Data Security Standard
PCI-SSC	Payment Card Industry- Security Standards Council Site
PME	Petites et Moyennes Entreprises
POC	Proof-Of-Concept
PrOnto	Privacy Ontology for legal reasoning
PSI	Politique(s) de Sécurité de l'Information
PS	Predicted Severity
QoS	Quality of Service
RDF	Resource Description Framework
RGPD	Règlement Général sur la Protection des Données
RMT	Risk Mitigation Times
SGBPAR	Severity Gap Between Predicted and Actual Risks
SGSI	Systèmes de Gestion de la Sécurité de l'Information
SLCT	Simple Logfile Clustering Tool
SME	Small and Medium-sized Enterprises
SMOG	Simple Measure of Gobbledygook
SoA	Statement of Applicability
SOC 2	Service Organization Controls 2
SOX	Sarbanes-Oxley Act
SPARQL	SPARQL Protocol and RDF Query Language
SWOT	Strengths, Weaknesses, Opportunities and Threats
TALN	Traitement Automatique de la Langue Naturelle
TCB	Total Compliance Budget
TF-IDF	Term Frequency-Inverse Document Frequency
TGÉC	Technologies de Gestion de l'Éthique et de la Conformité
TI	Technologies de l'Information
TO	Technologie Opérationnelle
UCO	Unified Cybersecurity Ontology
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VAC	Vérification Automatisée de la Conformité
VA	Valeurs Aberrantes
XML	Extensible Markup Language

LISTE DES ANNEXES

ANNEXE 1 : STRATÉGIE UTILISÉE POUR LA REVUE DE LITTÉRATURE . . .	117
ANNEXE 2 : STRUCTURE DE L'ONTOLOGIE	122

CHAPITRE 1 INTRODUCTION

La cybersécurité est devenue un enjeu majeur pour les entreprises et les organisations à l'ère du numérique. La prolifération des cybermenaces et l'évolution rapide des réglementations en matière de protection des données et de confidentialité rendent la gestion de la conformité en matière de cybersécurité complexe et cruciale. Dans ce contexte, il est essentiel pour les organisations de disposer d'outils et de méthodologies efficaces pour gérer les risques et assurer la conformité avec les exigences légales et réglementaires.

Ce premier chapitre sert d'introduction à notre étude sur la gestion de la conformité en matière de cybersécurité et la proposition d'un outil d'aide à la décision basé sur l'Intelligence Artificielle (IA). Nous commencerons par présenter la problématique liée à la cybersécurité et à la conformité, soulignant l'importance croissante de Gouvernance, Risque et Conformité (GRC) dans ce domaine. Ensuite, nous introduirons les concepts et les définitions clés qui serviront de base à notre recherche.

Par la suite, nous détaillerons les objectifs de cette recherche. Cette étude fait partie d'un partenariat de recherche et développement entre Polytechnique Montréal et une entreprise de services technologiques, mettant en évidence l'importance pratique et théorique de notre travail.

Enfin, nous présenterons le plan du mémoire, décrivant la structure et les thèmes abordés dans les chapitres suivants. Ce plan permettra d'orienter le lecteur à travers les différentes étapes de notre recherche, de la revue de la littérature à la méthodologie, en passant par la validation de l'ontologie et la conclusion générale.

1.1 Définitions des notions de base

1.1.1 Conformité et Conformité Informatique

La conformité est la pratique consistant à suivre les règles et réglementations établies par des personnes d'autorité [1]. L'étude [2] définit la conformité en cybersécurité comme l'état de conformité aux exigences de sécurité fonctionnelles imposées de l'extérieur et la fourniture de preuves de cette conformité.

En particulier, la conformité des Technologies de l'Information (TI) fait référence à la mise en œuvre et à l'exploitation de systèmes informatiques conformes à la législation applicable. Les lois et réglementations, les contrats avec les fournisseurs ou les consommateurs, les politiques internes de l'entreprise et les exigences non législatives telles que les normes et les meilleures pratiques sont tous des exemples d'exigences de conformité informatique [3].

1.1.2 Catalogues de contrôles et audit de sécurité

Les catalogues de contrôle sont des listes de pratiques de sécurité informatique conçues pour protéger les informations et les systèmes [4]. Le catalogue de contrôles de Federal Information Security Management Act (FISMA) définit un contrôle comme une sauvegarde ou une contre-mesure prescrite pour un système d'information afin de protéger la confidentialité, l'intégrité et la disponibilité de ses informations et de répondre à un ensemble d'exigences de sécurité définies [5].

Par ailleurs, l'audit de sécurité est le processus d'examen des enregistrements et des activités du système pour s'assurer qu'ils sont conformes à la politique et aux procédures de sécurité établies, en détectant toute faille dans les services de sécurité et en recommandant les modifications qui peuvent être nécessaires pour améliorer la sécurité [6].

1.1.3 Ontologies

L'ontologie est un cadre de connaissances partagées par des individus qui participent à un certain domaine [7]. Il peut être construit comme un compte-rendu sémantique formel du domaine avec des types d'idées et leurs relations rigoureusement décrites en langage logique, ou comme une structure conceptuelle informelle en langage simple. Une ontologie est une conceptualisation partielle d'un certain domaine de connaissance qui a été développée dans un langage formel et exploitable par une machine dans le but explicite de partager des informations sémantiques entre des systèmes automatisés dans l'environnement du Web sémantique. L'ontologie définit les concepts pertinents du domaine, leurs propriétés et les liens potentiels entre les concepts et les propriétés. Par conséquent, elle aide les systèmes automatisés à interpréter le contenu des ressources de manière significative, au lieu de se fier uniquement à leur structure physique ou à leurs éléments syntaxiques.

1.2 Problématique

La convergence croissante entre la cybersécurité et la conformité informatique soulève de nombreux défis pour les organisations, les professionnels et les chercheurs dans le domaine. Les entreprises doivent faire face à des réglementations de conformité de plus en plus complexes, tandis que l'automatisation de l'évaluation de la conformité et de l'attestation aux normes demeure un objectif difficile à atteindre. Parallèlement, les initiatives de gestion des connaissances rencontrent souvent des obstacles en raison de problèmes de mise en œuvre et d'intégration dans la culture organisationnelle. La problématique de cette recherche est donc de trouver des solutions efficaces pour combler l'écart entre la cybersécurité et la conformité informatique, tout en améliorant la gestion des connaissances et en automatisant les processus d'évaluation de la conformité.

Pour aborder cette problématique, nous devons d'abord comprendre les défis spécifiques auxquels les organisations sont confrontées lorsqu'elles tentent de se conformer aux réglementations et aux normes informatiques. Cela implique d'examiner les processus manuels actuels d'évaluation de la conformité et de certification des normes et d'identifier les obstacles à l'automatisation de ces processus. En outre, il est important de prendre en compte les défis liés à la préservation, au maintien et à l'amélioration de la mémoire et de l'expérience de l'entreprise.

Ensuite, il est crucial d'étudier comment les ontologies peuvent être utilisées pour faciliter la gestion des connaissances et l'automatisation des processus d'évaluation de la conformité. Les ontologies basées sur Web Ontology Language (OWL) offrent plusieurs avantages par rapport aux systèmes de gestion des connaissances traditionnels, tels que la capacité de décrire des structures de connaissances complexes et la prise en charge du raisonnement. Il est essentiel d'explorer comment ces avantages peuvent être utilisés pour résoudre les problèmes identifiés précédemment.

Enfin, il est nécessaire de mener des recherches pour combler le fossé entre la cybersécurité et la conformité informatique en développant des solutions intégrées qui traitent à la fois des aspects de sécurité et de conformité. Cela implique de prendre en compte les facteurs culturels, réglementaires, financiers et techniques qui contribuent aux problèmes de conformité et de cybersécurité, ainsi que d'examiner les conditions dans lesquelles la conformité présente des problèmes ayant un impact sur la cybersécurité.

En abordant ces questions, cette recherche vise à proposer une solutions innovante pour automatiser l'évaluation de la conformité et de l'attestation aux normes, améliorer la gestion des connaissances et combler l'écart entre la cybersécurité et la conformité informatique, permettant ainsi aux organisations de mieux protéger leurs données sensibles et de se conformer aux exigences réglementaires de manière plus efficace et efficiente.

1.3 Objectifs de la Recherche

1.3.1 Objectif Principal de la Recherche

L'objectif principal est de développer un outil automatisé d'aide à la décision en matière de gestion de la conformité informatique pour les Petites et Moyennes Entreprises (PME).

1.3.2 Sous-objectifs de la Recherche

- S-O1 : Déterminer les métriques existantes et les indicateurs clés de conformité pertinents pour la conformité informatique.
- S-O2 : Réaliser un outil permettant d'incorporer les règles et les contrôles de conformité dans une ontologie sémantique.
- S-O3 : Construire un pipeline d'apprentissage automatique pour prédire la conformité des journaux et des analyses de sécurité par rapport à un cas de test.
- S-O4 : Créer une preuve de concept (Proof-Of-Concept (POC)) pour la solution proposée, qui fournira des recommandations pour résoudre les problèmes de conformité informatique.

1.3.3 Finalités de l'étude et Intérêt pour le domaine

La nécessité de cette étude est née du fait que les PME ont du mal à se conformer à des réglementations complexes et en constante évolution et que le processus manuel de vérification de la conformité peut prendre du temps. Cette étude vise à apporter une solution à ces problèmes.

L'étude consiste à extraire des données de l'entreprise sous forme de journaux, qui seront ensuite analysées pour déterminer diverses caractéristiques du modèle d'affaires. Des métriques pertinentes seront établies en fonction des modèles de conformité utilisés, et ces métriques seront utilisées pour examiner l'impact des alertes et informer les responsables des actions futures nécessaires pour atteindre leurs objectifs commerciaux.

Les résultats attendus de cette étude sont : le développement d'un outil polyvalent qui fournirait des recommandations pour la conformité informatique, une définition de l'ontologie et des instructions pour l'appliquer avec l'IA. À long terme, l'objectif ultime de l'étude est de fournir un outil d'aide à la décision qui aiderait les cabinets d'audit à améliorer leur flux de travail et les entreprises auditées à obtenir un examen de conformité rapide et simplifié.

L'étude contribue également à l'avancement des connaissances sur la relation entre la conformité informatique, les ontologies et l'Apprentissage Automatique (AA). En combinant ces trois domaines d'expertise, cette recherche a le potentiel de fournir une perspective unique sur la façon d'évaluer et de vérifier efficacement la conformité informatique.

1.4 Définition du champ d'application

L'un des piliers de l'intelligence et du domaine de l'AA est l'organisation astucieuse des données en une matrice conceptuelle qui permet un raisonnement et une inférence automatiques basés sur les entités, les règles et les propriétés de celle-ci. Dans ce projet, nous avons choisi d'utiliser une ontologie pour conceptualiser les règles de sécurité d'un système d'information.

Un audit de sécurité est actuellement réalisé comme une série d'opérations visant à statuer sur la vulnérabilité des mesures¹ de sécurité selon des normes. Pour ce projet, nous limitons le champ d'application à un sous-ensemble des trois normes suivantes : ISO 27002 :2013 [9], PCI-DSS v3.2.1 [10] et NIST-SP 800-53 rev-5 [11]. En particulier, nous nous limitons au contrôle d'accès dans ISO 27002 :2013, les pare-feux dans PCI-DSS v3.2.1 et la gestion des comptes dans NIST-SP 800-53 rev-5.

Le choix de ces trois normes spécifiques repose sur plusieurs facteurs clés. Tout d'abord, l'ISO 27002 :2013 [9] est une norme internationalement reconnue qui fournit un ensemble complet de bonnes pratiques pour la gestion de la sécurité de l'information, couvrant un large éventail de domaines et d'aspects. La norme PCI-DSS v3.2.1 [10] est également considérée comme une référence en matière de sécurité pour les organisations qui traitent des données de cartes de paiement, en veillant à ce que ces données soient stockées, transmises et traitées de manière sécurisée. Enfin, le NIST-SP 800-53 rev-5 [11] est une norme émanant du gouvernement des États-Unis qui s'applique à tous les systèmes d'information fédéraux et

1. Une mesure de sécurité est un "moyen concret qui assure, partiellement ou totalement, la protection de l'actif informationnel contre une ou plusieurs menaces informatiques, et dont la mise en œuvre vise à amoindrir la probabilité de survenance de ces menaces ou à minimiser les pertes qui en résultent" [8].

qui fournit un ensemble de contrôles de sécurité exhaustifs et normalisés. La combinaison de ces trois normes garantit une couverture complète et diversifiée des meilleures pratiques en matière de sécurité, permettant ainsi d'aborder les différents aspects et défis liés à la sécurité des systèmes d'information. De plus, en se concentrant sur un sous-ensemble de ces normes, nous sommes en mesure de cibler efficacement les aspects les plus pertinents et applicables à notre projet, tout en assurant une approche structurée et cohérente pour l'audit de sécurité.

Par ailleurs, la décision de se concentrer uniquement sur le sous-ensemble traitant le contrôle d'accès, les pare-feux et la gestion des comptes s'explique par plusieurs raisons. Tout d'abord, étant donné les contraintes de temps et de ressources auxquelles nous sommes confrontés au cours de ce projet de recherche, il est essentiel d'adopter une approche ciblée et efficiente. Le contrôle d'accès, les pare-feux et la gestion des comptes sont des éléments clés de la sécurité de l'information, car ils jouent un rôle crucial dans la prévention des accès non autorisés et la protection des données sensibles. En nous concentrant sur ces aspects, nous abordons les questions de sécurité les plus critiques et les plus courantes auxquelles sont confrontées de nombreuses organisations, ce qui maximise l'impact et la pertinence de notre travail.

De plus, en choisissant d'examiner ces aspects spécifiques dans le cadre des trois normes sélectionnées (ISO 27002 :2013, PCI-DSS v3.2.1 et NIST-SP 800-53 rev-5), nous acquérons une compréhension approfondie des principes fondamentaux et des meilleures pratiques qui sous-tendent ces domaines clés de la sécurité de l'information. Cela permet d'extrapoler et de généraliser les résultats de notre recherche aux autres contrôles et parties de ces trois normes, car les mêmes principes et méthodologies peuvent être appliqués pour aborder les autres aspects de la sécurité. Bien que notre étude se limite à un sous-ensemble spécifique en raison des contraintes de temps et de ressources, les résultats obtenus sont suffisamment robustes et généralisables pour servir de base solide à des recherches futures et à des travaux de mise en œuvre dans d'autres domaines de la sécurité de l'information au sein de ces normes.

1.4.1 Sources de données et catégories

Les données utilisées dans ce projet sont de deux types :

- Type 1 : journaux² d'événements, journaux de sécurité et autres journaux qui sont générés automatiquement par les composants du réseau, les systèmes d'exploitation et les logiciels.

2. "journaux" et "logs" ont la même signification dans ce mémoire et sont utilisés de manière interchangeable

- Type 2 : données non structurées, légales, contractuelles ou de conformité issues de trois normes : ISO 27002 :2013 [9], PCI-DSS v3.2.1 [10] et NIST-SP 800-53 rev-5 [11], qui sont étiquetées à partir de l'ontologie conçue dans ce projet.

1.5 Considérations éthiques

L'étude n'a pas inclus d'expérimentation humaine directe et ne présente donc aucun risque pour les participants. Les chercheurs déclarent également qu'il n'existe aucun conflit d'intérêts. Cela permet de rappeler l'importance du respect des pratiques de recherche éthiques et démontre l'engagement des chercheurs à mener l'étude de manière responsable. Enfin, les chercheurs déclarent qu'ils n'ont aucun intérêt financier concurrent connu ni aucune relation personnelle qui pourrait sembler avoir influencé le travail rapporté dans ce mémoire.

1.6 Plan du mémoire

Ce mémoire est structuré en sept chapitres, présentant les différentes étapes de notre recherche et développement de la solution de gestion de conformité informatique pour les PME.

L'introduction de ce mémoire (le présent chapitre, i.e., Chapitre 1 : Introduction) décrit le contexte, les besoins, la problématique, les objectifs de recherche, les définitions des termes clés, le champ d'application et les considérations éthiques de notre étude. Elle est suivie d'une revue de la littérature (Cf. Chapitre 2 : Revue de la littérature), qui analyse les travaux existants et les connaissances actuelles sur la conformité informatique, la modélisation des connaissances et la cybersécurité. Cette revue permet de mieux comprendre les enjeux et les défis liés à la gestion de la conformité informatique pour les PME.

Le développement de la solution de gestion de conformité des TI est décrit en détail, en présentant la méthodologie de recherche utilisée, l'architecture globale de la solution et le Module de Collecte de Données (MCD), qui permet de rassembler les informations nécessaires pour évaluer la conformité (Cf. Chapitre 3 : Développement de la solution de gestion de conformité des TI). Ensuite, le Module de Vérification de Conformité (MVC) est abordé (Cf. Chapitre 4 : Module de Vérification de Conformité (MVC)), qui analyse les données collectées par le Module de Collecte de Données (MCD) pour évaluer la conformité des systèmes informatiques aux différentes normes et réglementations.

Dans le chapitre suivant (Cf. Chapitre 5 : Optimisation de l'interprétation et de la visualisation de la posture de conformité), l'Agent de l'Intelligence artificielle Orienté Sémantique (AIOS) est présenté, qui utilise la modélisation des connaissances pour faciliter l'évaluation de la conformité. Le Module de Tableaux de Bord (MTB) est également décrit, fournissant une interface utilisateur pour visualiser et interagir avec les résultats de l'évaluation de conformité.

Le sixième chapitre porte sur la validation de l'ontologie bonifiée (Cf. Chapitre 6 : Validation de l'ontologie bonifiée), un élément clé de notre solution qui permet d'assurer la pertinence et l'exactitude des connaissances modélisées pour la gestion de la conformité informatique. Enfin, une discussion des résultats obtenus, une évaluation des contributions de notre recherche et des perspectives pour les travaux futurs sont présentées (Cf. Chapitre 7 : Conclusion et Perspectives). L'étude se conclut en résumant l'ensemble du projet de recherche et en soulignant les principales réalisations.

CHAPITRE 2 REVUE DE LA LITTÉRATURE

Ce chapitre offre un résumé des principales conclusions issues de notre revue de littérature systématique, présentée en trois sections principales. La stratégie adoptée est détaillée dans l'Annexe 1. Dans un premier temps, nous introduirons le concept de conformité informatique et discuterons de son importance pour les PME. Nous aborderons les réglementations, normes et directives actuelles en matière de conformité informatique, ainsi que les défis liés à l'automatisation des processus d'évaluation et de certification. Par la suite, nous examinerons le rôle crucial de la modélisation des connaissances dans la préservation et l'amélioration de la mémoire organisationnelle et de l'expérience des entreprises. Enfin, la dernière section mettra en lumière les relations et les écarts entre la conformité informatique et la cybersécurité. En rassemblant les résultats des recherches existantes dans ces trois domaines, ce chapitre offre un aperçu éclairant de l'état actuel du domaine et jette les bases pour l'élaboration de la solution proposée. De plus, notre analyse de la littérature souligne les lacunes dans la recherche actuelle et identifie des directions potentielles pour les études futures.

2.1 Revue de la littérature sur les réglementations et les normes

Cette section se penche sur une multitude de facettes de la conformité informatique, incluant l'automatisation, les métriques, les stratégies de sécurité de l'information, ainsi que les éléments influençant la gestion de la conformité. Elle souligne également l'importance cruciale de la conformité, propose des méthodes pour son évaluation, et offre un aperçu du paysage actuel des solutions de conformité disponibles sur le marché.

2.1.1 Automatisation de la conformité

Automatisation de la conformité dans les deux domaines de protection de données et de la construction

Les organisations publiques et privées font face à des défis sans précédent pour assurer la conformité au Règlement Général sur la Protection des Données (RGPD) [12]. Des chercheurs et des technologues ont proposé des solutions d'ingénierie de la confidentialité pour aider les responsables à traiter les données de manière conforme et renforcer la confiance des personnes concernées. L'article [13] suggère d'utiliser le langage Prolog [14] pour modéliser un sous-ensemble du RGPD [12] et surveiller et vérifier la conformité des systèmes existants via leurs journaux. Cependant, l'article présente certaines limitations, notamment l'utilisation

de Prolog, qui souffre de restrictions pouvant entraver les développeurs de conformité.

Dans le domaine de la construction, l'article [15] explore diverses méthodes de Vérification Automatisée de la Conformité (VAC). Assurer la conformité dans la conception, la construction et l'utilisation d'un environnement bâti est essentiel pour garantir sa fonctionnalité, rentabilité, sécurité et durabilité. L'adoption du Building Information Modeling (BIM) [16] et de l'initiative Open Building Information Modeling (OpenBIM) [17] a facilité la normalisation des objets et attributs inclus dans les modèles de bâtiments, mais présente néanmoins des nouveaux défis, tels que la nécessité d'une formation adéquate du personnel pour maîtriser ces outils et la mise en place de systèmes de gestion de l'information efficaces. De plus, l'intégration de ces normes doit être réalisée de manière cohérente et harmonisée avec les processus de construction existants pour éviter d'éventuelles perturbations ou incompatibilités. Enfin, il convient de souligner l'importance d'une collaboration interdisciplinaire et d'un échange d'informations transparent entre les différents acteurs du secteur pour garantir l'efficacité et la réussite de ces méthodes de VAC.

La VAC [15] a progressé grâce aux efforts de chercheurs dans des domaines tels que l'OpenBIM basé sur les normes de représentation des connaissances et les solveurs de processus décrétés. Cependant, il reste des domaines nécessitant des recherches et développements supplémentaires, tels que l'amélioration de la qualité et l'exhaustivité des modèles BIM. Les méthodes de Traitement Automatique de la Langue Naturelle (TALN) (Nature Language Processing (NLP)) et d'IA offrent des perspectives prometteuses pour obtenir des traductions de meilleure qualité des codes et normes, mais les méthodes d'assurance et de contrôle de la qualité restent à explorer.

À l'avenir, la recherche se concentrera sur des systèmes tels que des serveurs de code calculable, des bibliothèques d'objets BIM, des processus de vérification normalisés et des plug-ins pour les outils BIM et d'autres outils similaires [15]. Ces développements permettront de progresser vers une automatisation plus robuste et fiable de la conformité dans divers domaines, dont la construction et la protection de données.

Automatisation de la conformité et évaluation de la sécurité dans les services infonuagiques

La croissance rapide des fournisseurs de services infonuagiques est largement due au modèle commercial de paiement à l'utilisation. Cependant, des obstacles importants freinent

l'adoption généralisée de l'infonuagique, tels que le manque de normes ouvertes et le coût élevé de leur mise en œuvre. De plus, certains fournisseurs manquent de motivation pour adopter des solutions ouvertes et conformes aux normes, préférant proposer des offres propriétaires et fermées. Le manque d'auditabilité en temps réel ou quasi-réel et l'absence de moyens pour comparer les fonctionnalités de sécurité des différents fournisseurs aux normes de l'industrie constituent également des défis. Pour surmonter ces problèmes, les auteurs de [18] ont développé un outil automatisé de conformité de la sécurité (en anglais, Automated Security Compliance Tool (ASCT)) qui vérifie et exprime la conformité en utilisant différentes techniques d'acquisition d'informations. Le prototype de preuve de concept de l'ASCT a été exposé à l'aide de l'Application Programming Interface (API) CloudAudit [19] et a été implémenté sur la plateforme infonuagique OpenStack [20].

L'article [18] s'appuie sur les travaux de la Cloud Security Alliance (CSA) [21], une organisation qui a élaboré plusieurs directives et cadres de référence, tels que CloudAudit [19] et Cloud Control Matrix (CCM) [22]. Ces ressources ont été conçues pour aider les fournisseurs de services infonuagiques à fournir des informations transparentes à leurs clients concernant la sécurité et la conformité de leurs offres. Cependant, des recherches supplémentaires sont nécessaires pour déterminer la meilleure façon de mettre en œuvre et d'intégrer efficacement ces outils dans les systèmes de gestion de la sécurité. L'outil de ASCT proposé par [18] peut être considéré comme une tentative prometteuse d'implémenter ces concepts et de renforcer la confiance des clients dans les services infonuagiques.

Trois autres initiatives prometteuses à considérer sont les travaux de [23], [24] et [25]. D'une part, Soren Bleikertz et ses collègues [23] ont développé une méthode automatisée pour évaluer la configuration de la sécurité réseau des plateformes infonuagiques utilisant l'Infrastructure-as-a-Service (IaaS). Cette méthode permet de récupérer les données de configuration et de prendre des décisions sans intervention humaine, mais se limite à l'évaluation de la sécurité du réseau. D'autres recherches se concentrent sur l'audit automatisé pour garantir la conformité à la sécurité infonuagique, comme l'illustre [26], qui examine les défis liés à l'audit en conformité avec la norme Payment Card Industry- Data Security Standard (PCI-DSS) dans un environnement infonuagique.

D'autre part, Piston Cloud Computing, Inc. [24] a conçu un framework pour automatiser la surveillance de la conformité infonuagique, en utilisant l'analyse des journaux pour collecter des informations. Cependant, cette approche présente des limitations, car toutes les informations nécessaires ne peuvent pas être obtenues uniquement par l'analyse des journaux, et

la communication interactive pour les contrôles de vulnérabilité est restreinte. En outre, le groupe de travail sur le protocole CloudTrust (CTP) [25] cherche à générer une confiance fondée sur des preuves, en s'assurant que tout ce qui se déroule dans l'infonuagique se passe réellement comme décrit.

Défis dans la mise en œuvre de l'automatisation

La construction d'une solution automatisée de conformité soulève plusieurs défis, comme l'a souligné [18]. Voici quelques-uns des obstacles majeurs qui doivent être surmontés pour réussir dans cette démarche :

- **Formalisation des exigences externes** : Il est crucial d'identifier et de formaliser l'ensemble des exigences externes que le système doit satisfaire pour garantir une sécurité adéquate et une conformité aux normes.
- **Manque de conseils de mise en œuvre** : Les normes de sécurité peuvent être complexes et manquer de conseils clairs sur la manière de les mettre en œuvre efficacement, ce qui rend difficile la création d'une solution automatisée.
- **Détermination des données nécessaires** : Il est important d'identifier les données ou les informations requises pour vérifier la conformité aux contrôles de sécurité, ce qui peut être un processus complexe et fastidieux.
- **Obtention des données des systèmes** : La collecte des données nécessaires à partir des systèmes peut être difficile en raison de restrictions d'accès, de problèmes d'interopérabilité ou de la complexité des systèmes.
- **Adaptation des systèmes existants** : Les systèmes existants (physiques ou infonuagiques) peuvent nécessiter des modifications pour faciliter l'extraction des données requises pour l'audit de conformité, ce qui peut être coûteux et nécessiter un investissement en temps et en ressources.
- **Sécurisation et intégrité des données d'audit** : Il est essentiel de protéger et de maintenir l'intégrité des données d'audit cruciales pour garantir la fiabilité des résultats de conformité et éviter les manipulations malveillantes.
- **Fonctionnement correct de l'outil automatisé** : Il est important de s'assurer que l'outil de conformité de sécurité automatisé fonctionne correctement et produit des résultats fiables pour garantir l'efficacité de la solution mise en place.

En plus de ces défis, il est également crucial d'assurer la mise à jour continue des outils automatisés de conformité pour répondre aux évolutions des normes de sécurité, des réglementations et des menaces. Une collaboration étroite entre les fournisseurs de services

informatiques, les experts en sécurité et les organismes de normalisation peut faciliter la mise en œuvre de solutions automatisées de conformité, en contribuant à l'élaboration de normes plus claires et à la mise en place de meilleures pratiques pour faciliter l'interopérabilité et la collecte de données.

2.1.2 Métriques, paramètres et indicateurs clés de la conformité

Les métriques de conformité en TI sont des indicateurs qui évaluent l'efficacité des programmes de conformité mis en œuvre [27]. Le non-respect de ces métriques peut entraîner des amendes et des coûts de remédiation. La mesure de la conformité peut être fastidieuse et des erreurs mineures peuvent causer des inexactitudes significatives. Ainsi, la conception de métriques de conformité doit permettre la collecte et le rapport de données fiables [27].

L'étude [28] rapporte certaines métriques de conformité applicables au domaine de l'informatique et à l'informatique en général :

1. Temps moyen entre l'identification et la résolution des problèmes de conformité externes.
2. Nombre de demandes de conseil ou d'assistance des employés pour des problèmes de conformité.
3. Nombre de violations de conformité non entièrement prouvées ou vérifiées.
4. Pourcentage d'opportunités d'amélioration de la conformité mises en œuvre avec succès.
5. Fréquence des revues de conformité (en jours).

Matt Kelly, fondateur de Radical Compliance, recommande cinq autres métriques de conformité importantes [27] :

1. Le délai moyen de découverte d'un problème (Mean Time for Incident Discovery (MTID)) mesure le temps nécessaire pour identifier un problème de conformité. Il est calculé en faisant la moyenne des temps de découverte des problèmes (Incident Discovery Time (IDT)) pour un nombre d'incidents donné (n) (voir l'équation 2.1).
2. Le temps moyen de résolution des problèmes (Mean Time for Incident Resolution (MTIR)) quantifie le temps nécessaire pour résoudre un problème identifié. Il est obtenu en faisant la moyenne du temps total pour résoudre tous les problèmes (Incident Resolution Time (IRT)) pour un nombre de problèmes donné (n) (voir l'équation 2.2).

3. La métrique des dépenses de conformité par problème (Compliance Expense Per Incident (CEPI)) est le rapport du budget total de conformité (Total Compliance Budget (TCB)) au nombre de problèmes de conformité (Number of Compliance Incidents (NCI)). Elle est calculée trimestriellement ou annuellement (voir l'équation 2.3).
4. L'écart de gravité entre les risques prédits et réels (Severity Gap Between Predicted and Actual Risks (SGBPAR)) quantifie la différence entre la gravité prédite (Predicted Severity (PS)) d'un risque et sa gravité réelle (Actual Severity (AcS)) aux niveaux financier et opérationnel (voir l'équation 2.4).
5. Le temps moyen d'atténuation des risques (Mitigation Mean Time (MMT)) est mesuré en divisant la somme de tous les temps d'atténuation des risques (Risk Mitigation Times (RMT)) par le nombre total de risques surveillés (n). Cette métrique montre la rapidité avec laquelle le programme de conformité s'adapte aux risques changeants (voir l'équation 2.5).

$$MTID = \frac{1}{n} \sum_{i=1}^n IDT \quad (2.1)$$

$$MTIR = \frac{1}{n} \sum_{i=1}^n IRT \quad (2.2)$$

$$CEPI = \frac{TCB}{NCI} \quad (2.3)$$

$$SGBPAR = |PS - AcS| \quad (2.4)$$

$$MMT = \frac{1}{n} \sum_{i=1}^n RMT \quad (2.5)$$

Ces métriques offrent une vue d'ensemble de l'efficacité d'un programme de conformité et permettent aux entreprises d'identifier les domaines qui nécessitent des améliorations. En utilisant ces métriques, les organisations peuvent surveiller et évaluer leur performance en matière de conformité, ainsi qu'identifier les problèmes potentiels et les résoudre de manière proactive.

Pour tirer le meilleur parti de ces métriques, il est essentiel de les intégrer dans un système de gestion de conformité bien conçu. Un tel système doit inclure des processus pour collecter

et analyser les données, ainsi que des mécanismes pour communiquer les résultats aux parties prenantes concernées. De plus, les organisations doivent régulièrement revoir et ajuster leurs métriques de conformité afin de s'assurer qu'elles restent pertinentes et efficaces face aux évolutions technologiques et réglementaires.

2.1.3 Conformité informatique et Politiques de Sécurité de l'Information (PSI)

La conformité informatique et les Politique(s) de Sécurité de l'Information (PSI) sont des sujets de recherche populaires parmi les chercheurs en cybersécurité et représentent également une question de gestion cruciale pour les organisations [29]. Plusieurs études se concentrent sur les facteurs qui déterminent le comportement de non-conformité des employés aux PSI [30, 31]. Les organisations mettent en œuvre des PSI pour améliorer l'efficacité des contrôles de cybersécurité, gérer les impacts commerciaux, financiers et juridiques, et structurer la conformité aux normes et réglementations nationales et internationales [32–34]. Un défi majeur réside dans la manière d'encourager la conformité avec les PSI [32, 35].

L'approche traditionnelle de la conformité, axée sur l'utilisation de forces externes, s'appuie sur des mécanismes de contrôle coercitifs et des incitations monétaires pour encourager le comportement de conformité souhaité [36]. Cependant, la gestion de la cybersécurité et la maximisation de la conformité peuvent être plus efficaces si l'accent est mis sur des aspects non techniques, tels que le développement d'une culture de la conformité et la création d'un climat éthique au sein de l'organisation [32, 37, 38]. Les universitaires s'accordent sur le fait que la conformité réglementaire améliore les performances en matière de cybersécurité et la responsabilité des mécanismes et procédures de cybersécurité [32, 39].

Les considérations financières jouent un rôle important dans la réalisation des objectifs de conformité organisationnelle [32]. Les besoins financiers des organisations pour les activités de cybersécurité, tels que l'amélioration de l'infrastructure de cybersécurité et la mise en place de programmes de sensibilisation, d'éducation et de formation, nécessitent l'allocation de ressources humaines et financières importantes. Les organisations cherchent à optimiser la gestion de leurs ressources limitées tout en améliorant leur posture de conformité informatique [32]. Ainsi, il est essentiel de trouver un équilibre entre les besoins financiers, la gestion des risques et l'atteinte des objectifs de conformité.

2.1.4 Facteurs affectant la gestion de la conformité

Plusieurs facteurs influencent la gestion de la conformité dans les organisations, y compris la formation, l'éducation et la sensibilisation aux PSI, ainsi que les moyens de dissuasion mis en place [40]. La recherche montre que la gravité et l'efficacité de la menace perçue sont les prédicteurs les plus forts des intentions de conformité et du comportement des employés concernant l'utilisation d'équipements personnels à des fins professionnelles (Bring Your Own Device (BYOD)) [41]. Cependant, la gravité de la menace perçue n'a pas eu d'effet significatif sur l'intention comportementale des employés aux États-Unis, suggérant un manque de sensibilisation à l'utilisation abusive des téléphones intelligents [42].

L'intégration de facteurs sociaux, techniques, environnementaux et organisationnels doit être prise en compte en raison de la nature de l'utilisation des appareils mobiles [43]. Les gestionnaires devraient également s'interroger sur les facteurs environnementaux qui poussent les employés à adopter des comportements négatifs en matière de cybersécurité [44]. Le renforcement des comportements souhaitables peut produire de meilleurs résultats [45].

L'optimisation de la gestion des ressources et l'amélioration de la posture de conformité informatique, mentionnées précédemment, nécessitent donc une approche holistique tenant compte de ces facteurs. Les organisations doivent non seulement investir dans la formation, l'éducation et la sensibilisation, mais aussi créer un environnement qui encourage les comportements positifs en matière de cybersécurité. En outre, elles doivent mettre en place des mécanismes pour dissuader les employés d'adopter des comportements négatifs et s'assurer que les menaces sont clairement perçues pour renforcer les intentions de conformité.

2.1.5 Aperçu du marché des solutions de conformité

Cette section offre un aperçu du marché des solutions de conformité axées sur la cybersécurité et l'informatique. Le marché des produits de conformité est vaste et couvre divers domaines tels que la gouvernance informatique, les risques liés à l'information, la protection de la vie privée, le droit, la finance et la gestion des marques. La Figure 2.1 illustre l'analyse de Gartner pour 2021 du marché des produits de conformité [46].

Selon cette analyse de Gartner, les éléments clés à retenir sont les suivants :

1. Gouvernance Dynamique des Risques (GDR) : La GDR assure une gestion des risques alignée sur la stratégie de l'organisation en adaptant dynamiquement la gouvernance des risques en fonction de chaque situation.

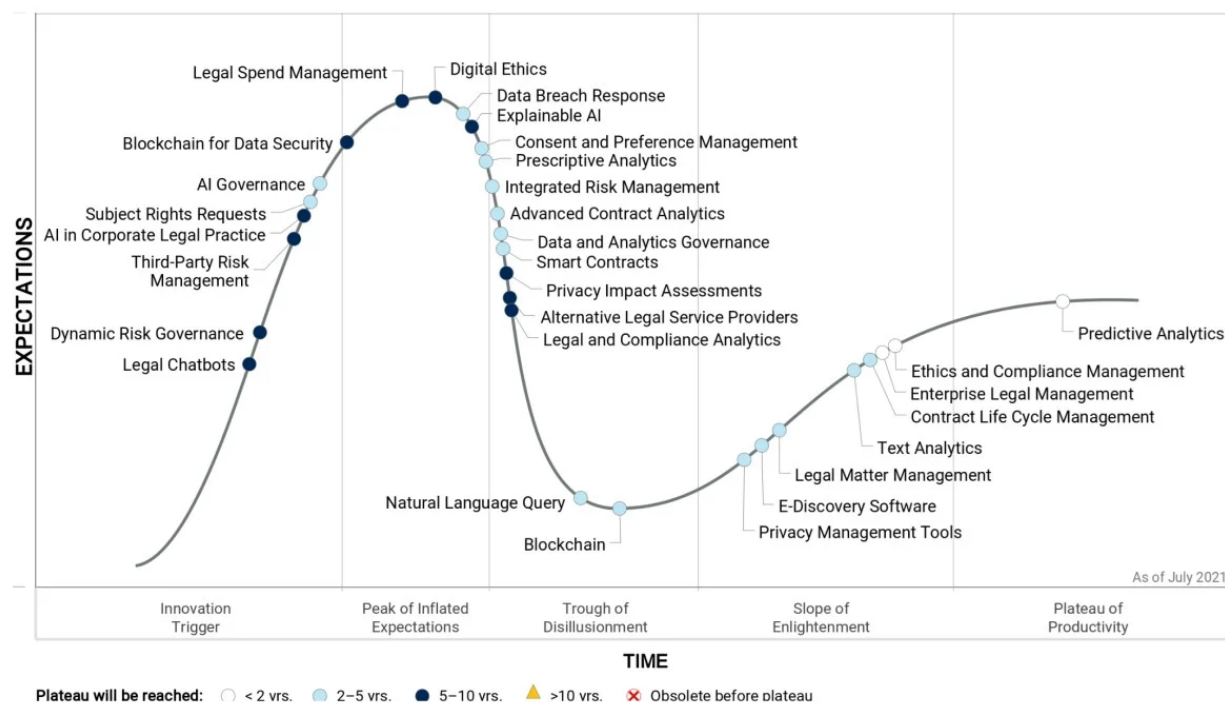


FIGURE 2.1 Cycle de Hype Gartner pour les technologies juridiques et de conformité [46]

2. Gestion des Risques Liés aux Tiers (GRLT) : Ce composant concerne la gestion des risques associés aux parties prenantes externes, tels que les fournisseurs, distributeurs, agents, partenaires. La GRLT couvre divers processus métier de gestion des risques, y compris les risques juridiques, de conformité, de confidentialité, informatiques et de chaîne d'approvisionnement.
3. Gestion Intégrée des Risques (GIR) : La GIR englobe la technologie, les processus d'affaires et les données pour simplifier, automatiser et intégrer divers domaines de risques au sein d'une organisation. Elle facilite la visibilité des fonctions clés de risque et de conformité, ainsi que des partenaires commerciaux critiques, des fournisseurs et des entités externalisées.
4. Analyse Avancée des Contrats (AAC) : L'AAC utilise le TALN et l'IA pour analyser les obligations contractuelles d'une organisation, identifier les problèmes potentiels et recommander les actions appropriées.
5. Évaluations de l'Impact sur la Vie Privée (EIVP) : Les EIVP aident les organisations à identifier, évaluer et gérer les risques liés à la vie privée et au traitement des informations personnelles.
6. Analyse Juridique et de Conformité (AJC) : L'AJC emploie des analyses descriptives,

prescriptives et prédictives des processus juridiques et de conformité pour améliorer la gestion des risques, optimiser les coûts et augmenter la productivité des équipes travaillant dans ces domaines.

7. Gestion du Cycle de Vie des Contrats (GCVC) : La GCVC est utilisée pour gérer les contrats, de l'initiation à la négociation, la signature électronique, les amendements, et le renouvellement ou la résiliation éventuelle des accords contractuels.
8. Technologies de Gestion de l'Éthique et de la Conformité (TGÉC) : Les TGÉC visent à optimiser les processus décisionnels des acteurs de la conformité dans la gestion des risques de conformité et d'éthique.

En somme, l'analyse de Gartner met en lumière les diverses solutions de conformité disponibles sur le marché, qui peuvent aider les organisations à gérer les risques et à respecter les réglementations dans un environnement informatique et de cybersécurité en constante évolution. En tenant compte des besoins en matière de conformité informatique et des facteurs qui influencent la gestion de la conformité, les organisations peuvent sélectionner et mettre en œuvre les solutions les plus adaptées pour améliorer leur posture de sécurité globale.

2.2 Revue de la littérature sur la modélisation des connaissances

La modélisation des connaissances est une composante essentielle de la recherche et du développement en matière de cybersécurité et de conformité. Elle permet de représenter, de structurer et de partager les informations et les connaissances acquises dans ce domaine, facilitant ainsi l'interopérabilité, l'automatisation et l'amélioration continue des processus et des pratiques de sécurité. Dans cette section, nous passons en revue la littérature sur la modélisation des connaissances en mettant l'accent sur les classifications de la connaissance, le Web Ontology Language (OWL) [47] et l'utilisation et la pertinence des ontologies pour la cybersécurité et la conformité. Les trois paragraphes 2.2.1, 2.2.2 et 2.2.3 sont basés sur les travaux de [48–50].

2.2.1 Classifications de la connaissance

La connaissance est un concept multidimensionnel étudié dans diverses disciplines, comme la sociologie, la psychologie, l'IA et la philosophie. Elle est généralement définie comme la compréhension ou la familiarité avec des informations, des faits ou des compétences acquises par l'expérience ou l'éducation [51]. Pourtant, la classification des types de connaissances soulève des questions sur leur pertinence et leur clarté.

La distinction entre connaissances déclaratives et procédurales est couramment utilisée pour classer les connaissances, mais d'autres classifications existent [51]. Les connaissances déclaratives concernent des faits ou informations spécifiques, tandis que les connaissances procédurales portent sur l'exécution d'une tâche. Une autre approche de classification oppose les connaissances explicites, facilement exprimées et transférées, aux connaissances implicites, difficiles à articuler et souvent inconscientes.

La pertinence et la clarté de ces classifications sont sujettes à débat. Par exemple, la distinction entre connaissances explicites et implicites peut être floue, car certaines connaissances peuvent être transférées dans une certaine mesure mais restent difficiles à exprimer complètement. De plus, ces classifications pourraient ne pas être suffisamment exhaustives pour couvrir toutes les formes de connaissances existantes.

Dans les domaines de l'informatique et de l'IA, les ontologies sémantiques sont souvent utilisées pour représenter et organiser les connaissances de manière structurée et formelle [52]. Elles offrent un moyen de modéliser les relations entre différents concepts et entités et permettent de raisonner et tirer des conclusions sur ces informations. Cependant, les ontologies sémantiques se concentrent principalement sur les connaissances déclaratives, ce qui limite leur applicabilité à d'autres types de connaissances, comme les connaissances procédurales ou implicites.

En dernier lieu, il est essentiel de reconnaître les limites des classifications de la connaissance pour éviter de simplifier excessivement la compréhension de ce concept complexe. Cette analyse critique des classifications de la connaissance vise à contribuer à une étude de recherche plus nuancée et pertinente.

2.2.2 Web Ontology Language (OWL)

Le Web Ontology Language (OWL) [47] joue un rôle crucial dans la représentation des concepts qui forment les ontologies, permettant l'accumulation de connaissances sur des modèles abstraits appelés instances de classes ou individus. L'efficacité de cette accumulation dépend en grande partie des relations définies par le concepteur de l'ontologie, qui servent à lier les classes ou les individus. OWL se distingue en utilisant des propriétés d'objet pour représenter ces relations de manière personnalisée et directionnelle.

Cependant, pour augmenter la précision de la représentation, OWL ne se contente pas de définir simplement ces relations. Il approfondit leur description en utilisant des caractéristiques de propriété et des restrictions. Les premières peuvent être appliquées aux propriétés d'objet, aux propriétés d'annotation et aux propriétés de type de données, tandis que les secondes, comprenant des catégories existentielles, universelles, has-value et cardinales, ajoutent des contraintes spécifiques sur les relations.

Évaluer le succès d'une ontologie nécessite l'utilisation de métriques ontologiques spécifiques, comme l'expressivité, la sémantique et le niveau de rigueur mathématique. Cependant, ces métriques présentent des défis, comme la subjectivité inhérente à l'évaluation de la sémantique, et la difficulté de déterminer précisément si une ontologie est "légère" ou "lourde", une distinction qui dépend de la rigueur mathématique.

C'est ici qu'intervient Owlready2 [49], un outil Python recommandé pour gérer les ontologies OWL. Son attrait réside dans sa syntaxe de haut niveau, qui simplifie l'accès aux entités d'une ontologie. Cet outil pourrait être d'une grande aide pour automatiser l'entrée des éléments de conformité et des contrôles, et pour tester l'ontologie de conformité proposée dans ce mémoire, rendant ainsi le processus d'évaluation plus précis et efficace.

Malgré ces outils et méthodes, la validation d'une ontologie reste un processus complexe. Comme suggéré par [53], elle se déroule généralement en deux étapes : évaluation de l'ontologie en tant qu'outil d'interrogation et comparaison des résultats d'extraction de connaissances avec les processus humains. Cependant, cette méthode pourrait ne pas être suffisamment rigoureuse pour toutes les applications d'ontologie, soulignant la nécessité d'une approche de validation plus systématique et personnalisée.

2.2.3 Utilisation et pertinence des ontologies

Les ontologies sémantiques sont de plus en plus employées dans le domaine des TI et de la sécurité, trouvant des applications dans divers contextes [54]. En raison de leur capacité à représenter et organiser les informations de manière structurée et formelle, elles sont largement adoptées dans des domaines tels que la gestion des connaissances, le Traitement Automatique de la Langue Naturelle (TALN) et l'Intelligence Artificielle (IA).

Dans le domaine de la cybersécurité, les ontologies sémantiques sont utilisées pour des tâches telles que la détection des intrusions, le renseignement sur les menaces et la réponse

aux incidents [55]. Cependant, il est important de souligner que leur efficacité dans l'automatisation et l'amélioration du processus de détection et de réponse aux incidents de sécurité dépend fortement de la qualité et de la pertinence de l'ontologie développée. Une ontologie bien conçue peut fournir une représentation formelle et structurée des connaissances liées à la cybersécurité, y compris les menaces, les vulnérabilités et les incidents.

Les ontologies sémantiques sont également appliquées à la représentation des PSI et des exigences de conformité [56]. Un cadre commun basé sur les ontologies sémantiques peut faciliter la compréhension et l'application des exigences de sécurité et de conformité des informations. En modélisant les relations entre les politiques de sécurité, les réglementations et les systèmes et actifs informatiques concernés, les ontologies peuvent aider les organisations à garantir la conformité de leurs systèmes informatiques aux normes, aux réglementations et aux meilleures pratiques du secteur. Toutefois, il convient de noter que les ontologies ne sont pas une solution miracle et que leur efficacité dépendra de leur adéquation aux besoins spécifiques de l'entreprise et de leur mise à jour régulière pour tenir compte des évolutions technologiques et réglementaires.

De nombreux projets de recherche sur la cybersécurité utilisent des ontologies. Ils ont examiné des concepts de haut niveau, mais aussi des aspects spécifiques du domaine [57–60]. Ces ontologies décrivent certains aspects du domaine, tels que les réseaux, les vulnérabilités, les vecteurs d'attaque, les stratégies de défense, les actifs, les protocoles et les outils de sécurité [61]. Dans ce projet, plusieurs ontologies existantes ont été identifiées et utilisées pour alimenter l'ontologie de conformité envisagée. La liste suivante présente un échantillon pertinent d'ontologies et de graphes de connaissances identifiés dans la littérature :

- **Unified Cybersecurity Ontology (UCO)** [62] : L'ontologie UCO est destinée à prendre en charge l'intégration des informations et de la connaissance de la situation cybernétique dans les systèmes de cybersécurité.
- **D3fend** [63] : Un graphe de connaissances sur les contre-mesures en cybersécurité.
- **Ontology-driven Threat Modeling (OdTM) Framework** [64, 65] : est une ontologie pour la modélisation automatique des menaces des systèmes informatiques.
- **Insider Threat Indicator Ontology** [66] : Une ontologie conçue pour prendre en charge la détection, la création, le partage et l'analyse des indicateurs de menace d'initié.
- **IT Service Management Ontology (ITSMO)** [67] : Ontologie pour la gestion des services de TI
- **Adaptive Mapping of Security Standard ontology (AMSS)** [59] : Prototype

- de visualisation des informations de mappage entre différentes normes de sécurité
- **DeTT&CT** [68] : DeTT&CT vise à aider les équipes de cyberdéfense à utiliser le cadre ATT&CK de MITRE pour évaluer et comparer la qualité des sources de données, la couverture de visibilité, la couverture de détection et les comportements des acteurs de la menace.
 - **Financial Industry Business Ontology (FIBO)** [69] : FIBO est une ontologie formelle qui fournit un vocabulaire commun pour les contrats financiers et les concepts connexes.
 - **Legal Knowledge Interchange Format (LKIF)** [70] : LKIF est une bibliothèque d'ontologies pertinentes pour le domaine juridique.
 - **MITRE Att&ack Navigator** [71] : ATT&CK Navigator est un outil Web permettant d'annoter et d'explorer les matrices du cadre ATT&CK de MITRE
 - **NEURONA** [72] : Ontologie pour la protection des données personnelles et le droit
 - **GDPR-Onto** [73] : fournit des ontologies pour en relation avec le RGPD [12]
 - **Privacy Ontology for legal reasoning (PrOnto)** [74] : PrOnto est une ontologie qui modélise les composants conceptuels fondamentaux du RGPD. PrOnto a été créée à l'aide d'une technique interdisciplinaire appelée Methodology for building Legal Ontology (MeLOn) [75]. MeLOn est spécifiquement développée pour les ontologies juridiques et les problèmes que rencontrent les opérateurs juridiques lorsqu'ils utilisent des techniques ontologiques, des processus de conception de patrons ou des approches sémantiques pour définir un modèle de la réalité.
 - **Ontology for a HIPAA Compliant Cloud Service (OntoHCCS)** [76] : Une ontologie qui décrit les règles de confidentialité et de sécurité de l'Health Insurance Portability and Accountability Act (HIPAA) [76].
 - **L'ontologie GConsent** [77] : Une ontologie qui permet de modéliser et traiter le concept de consentement dans le RGPD.

Dans l'ensemble, les travaux existants sur les ontologies en cybersécurité et conformité ont jeté les bases pour la modélisation des connaissances et la représentation des concepts dans ce domaine. Cependant, il reste des défis à relever pour combler les lacunes et améliorer l'automatisation, la réutilisation des connaissances, et la couverture de l'évaluation de la conformité. Les futures recherches devraient se concentrer sur le développement d'ontologies plus complètes et intégrées qui couvrent un large éventail d'aspects de la cybersécurité et de la conformité, tout en facilitant l'interopérabilité et la réutilisation des connaissances entre différents domaines et contextes.

2.3 Illusions des normes de conformité informatique

Le domaine de la conformité informatique rencontre un ensemble complexe de défis et de problèmes. Ces difficultés se manifestent dans divers sous-domaines et contextes. Par exemple, l'illusion de la conformité peut être présente dans de nombreux aspects de la sécurité informatique, où une organisation peut croire à tort qu'elle est conforme. Cette section synthétise la littérature pertinente concernant les dangers potentiels de ces illusions et les stratégies d'adaptation des processus de conformité informatique pour prévenir ou atténuer ces risques.

2.3.1 Défis et perspectives dans la conformité informatique

Les approches de conformité informatique actuelles rencontrent des difficultés majeures lorsqu'il s'agit d'assurer une sécurité optimale, notamment en raison du décalage entre les recommandations de haut niveau des normes spécifiques et les informations de journalisation de bas niveau disponibles dans les infrastructures infonuagiques [78]. Ce décalage complique l'interprétation des règles de sécurité et la définition des propriétés de sécurité utilisables dans les solutions d'audit. De plus, les méthodes actuelles reposant sur l'identification manuelle des attributs de sécurité sont sujettes aux erreurs et difficilement réalisables dans un environnement dynamique tel que l'infonuagique, où l'automatisation du processus d'audit est cruciale pour assurer précision et efficacité [78].

L'utilité des normes informatiques est également remise en question, car elles peuvent être trompeuses en associant des scores d'audit de conformité élevés à une sécurité renforcée [79]. Bien qu'elles puissent servir de lignes directrices générales, elles deviennent problématiques lorsqu'elles sont considérées comme des exigences légales de la liste de contrôle. Une étude en deux parties [79] a révélé que le respect rigoureux des programmes de conformité dans plusieurs normes telles que l'Internal Revenue Service Publication 1075 (IRS P1075) [80], la PCI-DSS et la North American Electric Reliability Corporation Critical Infrastructure Protection (NERC CIP) [81] a conduit à la découverte de 148 problèmes de sécurité, dont certains pourraient être résolus avec des révisions mineures des normes et d'autres nécessiteraient des investissements importants. Cette étude met en lumière les défis liés à l'élaboration de normes adaptées à l'évolution rapide du paysage de la sécurité, suffisamment complètes pour être appliquées dans diverses situations et suffisamment robustes pour être utilisées comme listes de contrôle d'audit de conformité. Cependant, aucune des normes examinées ne dispose de procédures pour signaler les problèmes de sécurité, ce qui est préoccupant [79].

Plusieurs autres études ont mis en évidence les limites de la conformité informatique en tant que garantie de sécurité. L'étude de [82] a révélé que même les organisations conformes à l'ISO 27001 peuvent présenter des vulnérabilités significatives. De même, l'article de [83] a montré que la conformité à la PCI-DSS ne garantit pas nécessairement la protection des données sensibles des clients si des contrôles de sécurité importants sont négligés.

Face à ces constats, les organisations doivent adopter une approche proactive pour identifier et atténuer les risques potentiels, même lorsqu'elles sont conformes à une norme spécifique. Cela implique une approche personnalisée de la conformité informatique, adaptée aux besoins et aux circonstances particulières de l'organisation, et visant à anticiper et à gérer les risques de sécurité.

2.3.2 Regroupement et analyse des journaux

Le regroupement¹ et l'analyse des journaux de sécurité peuvent être considérées comme une bonne méthodologie pour vérifier et automatiser la conformité aux normes informatiques pour plusieurs raisons [84-87] :

1. Axé sur les données : l'analyse des journaux de sécurité donne un aperçu de la façon dont les systèmes et les réseaux se comportent et sont utilisés. Cette approche basée sur les données permet d'identifier les risques potentiels, les failles de sécurité et les écarts par rapport aux normes informatiques établies.
2. Amélioration de la sécurité : En analysant les journaux de sécurité, nous pouvons identifier les modèles et les anomalies qui peuvent indiquer une violation de la sécurité. Cela permet de répondre rapidement aux menaces potentielles de sécurité et d'améliorer la sécurité globale de systèmes.
3. Meilleure visibilité : Le regroupement et l'analyse des journaux peuvent fournir une meilleure visibilité sur les opérations de systèmes informatiques, ce qui peut aider à identifier les problèmes potentiels avant qu'ils ne deviennent des problèmes majeurs. Cela peut également permettre une compréhension plus complète du paysage informatique global.

Il existe de nombreuses façons de regrouper les journaux dans la cybersécurité, mais il peut être difficile de choisir le meilleur outil pour un travail spécifique. En effet, les objectifs, hypothèses et caractéristiques des journaux varient selon les domaines. Pour résoudre ce

1. Regroupement est la traduction française utilisée dans ce mémoire pour le mot anglais "clustering"

problème, les chercheurs dans [88–93] ont effectué une revue de la littérature sur les approches de regroupement, en les catégorisant en fonction de leurs attributs pertinents, tels que les objectifs et les techniques de regroupement. Ils ont ensuite créé un ensemble de critères d'évaluation à utiliser lors de l'évaluation de 65 approches différentes et de 2 solutions non académiques.

L'enquête [88–93] a permis de découvrir que la plupart des approches de regroupement visent à atteindre un ou plusieurs des quatre objectifs principaux suivants : vue d'ensemble et filtrage, analyse syntaxique et extraction de signatures, détection des aberrations statiques, et séquences et détection dynamique des anomalies. De plus, les auteurs ont découvert que les techniques de regroupement utilisées dans ces approches appartenaient à différentes catégories, telles que le partitionnement, les réseaux neuronaux, l'analyse du code source, les techniques basées sur les jetons, sur les caractères, sur la distance et sur la densité. Enfin, l'étude a permis de déterminer quelles approches sont les mieux adaptées à la détection de types spécifiques d'anomalies dans les données de journaux et d'évaluer les méthodes utilisées pour évaluer les algorithmes de regroupement.

Sur la base des découvertes de cette enquête, nous soutenons qu'un outil comme Logdata Anomaly Miner (AMiner) [88–93] est le meilleur choix pour effectuer le regroupement de journaux sur l'état de conformité par rapport à des contrôles spécifiques. AMiner est un outil hautement flexible et configurable qui est conçu pour fonctionner avec des ressources limitées et avec les permissions les plus faibles possibles, ce qui le rend idéal pour une utilisation dans des environnements de serveurs de production. AMiner a été validé et soutenu par de nombreux articles de recherche et brevets industriels [94]. En outre, la capacité d'AMiner à définir des pipelines d'analyse pour la détection des anomalies et à analyser les données des journaux en fait une solution idéale pour ceux qui cherchent à mettre en œuvre des approches de liste blanche basées sur le comportement et l'identification des activités légitimes. En fin de compte, cela fait d'AMiner une option très intéressante pour ceux qui cherchent à effectuer le regroupement de journaux et la détection d'anomalies dans le domaine de la cybersécurité en général et la conformité en particulier.

2.4 Discussion

Cette revue de littérature a permis de mettre en évidence un certain nombre de lacunes et de défis dans la recherche actuelle sur l'automatisation de la conformité et l'utilisation des ontologies en cybersécurité. Tout d'abord, malgré la reconnaissance de l'importance des

ontologies pour la représentation des connaissances, leur adoption et leur utilisation restent limitées, en raison de la complexité inhérente à leur création et à leur maintenance [47]. De plus, l'incorporation de techniques d'AA dans la modélisation des connaissances en matière de cybersécurité et de conformité est encore peu explorée, malgré son potentiel considérable pour améliorer l'efficacité et la précision des évaluations de conformité.

Par ailleurs, les approches actuelles pour mesurer et évaluer la conformité informatique peuvent être limitées ou simplistes. Il serait bénéfique d'examiner des approches plus holistiques qui prennent en compte des aspects à la fois qualitatifs et quantitatifs. Pour ce faire, nous proposons d'investir du temps et des ressources dans la création et la maintenance d'une ontologie adaptée à notre contexte de conformité, en utilisant des outils tels que Owlready2 [49] pour faciliter l'accès aux entités d'une ontologie.

La gestion de la conformité est influencée par un éventail de facteurs, y compris les différences culturelles, organisationnelles et technologiques. Ces facteurs peuvent influencer la façon dont les questions de conformité sont perçues et gérées dans différents contextes [53]. Par conséquent, pour garantir une compréhension commune et une intégration réussie des différentes approches, nous encourageons la collaboration et le partage des connaissances entre les experts en cybersécurité, les développeurs d'ontologies et les chercheurs en AA.

Enfin, la recherche actuelle pourrait ne pas couvrir exhaustivement le marché des solutions de conformité en cybersécurité et informatique. Cela soulève plusieurs questions de recherche, notamment sur la viabilité à long terme de la sécurité obtenue par la mise en œuvre de normes de conformité, la possibilité de garantir une conformité complète, et la mesure de l'incertitude dans une évaluation de la conformité.

Pour combler ces lacunes, notre étude de recherche adoptera une approche holistique pour aborder la conformité informatique. Cette approche inclut l'examen des normes, des cadres normatifs, des systèmes de gouvernance, des métriques de conformité et des facteurs humains, comportementaux et stratégiques. Notre étude s'appuiera sur les meilleures pratiques et les enseignements tirés de la littérature existante pour aborder les problématiques complexes de la conformité informatique, tout en explorant des solutions intégrées et automatisées pour améliorer la cybersécurité et la conformité.

CHAPITRE 3 DÉVELOPPEMENT DE LA SOLUTION DE GESTION DE CONFORMITÉ DES TI

Ce chapitre présente les diverses méthodes de recherche employées au cours de cette étude. Une approche mixte a été mise en œuvre, combinant des méthodes qualitatives et quantitatives pour recueillir et analyser les données. Les sections suivantes offrent un aperçu approfondi de la méthodologie de recherche, de l'architecture de la solution proposée et du Module de Collecte de Données (MCD).

3.1 Méthodologie de recherche

La méthodologie Action Design Research (ADR) [95] est utilisée dans ce projet. Nous l'avons choisi pour notre recherche parce qu'elle est particulièrement adaptée à la conception et à l'évaluation de solutions pratiques à des problèmes complexes. Action Design Research (ADR) met en œuvre l'approche Design Science Research (DSR) proposée par [96], qui est basée sur un paradigme de résolution de problèmes afin de générer des artefacts pour l'ingénierie et les sciences appliquées. L'ADR fournit le processus pour réaliser la construction d'artefacts, de modèles et de méthodes appliqués au développement et à l'utilisation de systèmes d'information basés sur l'approche Design Science Research (DSR) ([96–99]). Les artefacts créés et évalués par un processus itératif axé sur la compréhension de l'expérience vécue des parties prenantes réussissent à fournir des connaissances dans les projets informatiques et de cybersécurité ([95, 100, 101]).

Dans ce projet de recherche, ADR est utilisée pour concevoir et mettre en œuvre une Preuve de Concept pour un système qui peut aider les organisations à gérer leurs problèmes de conformité informatique. Le cadre théorique utilisé par ADR se concentre sur la création d'artefacts, les livrables du projet, à travers le cycle de Construction, Mise en Oeuvre et Évaluation (CMÉ). Ainsi, alors qu'un artefact est attendu, voire anticipé, à la fin du projet, plusieurs artefacts intermédiaires seront produits, mis en œuvre et évalués au cours du projet. Le cycle CMÉ tient compte des intentions du chercheur, des besoins exprimés et de l'utilisation continue des artefacts dans un contexte particulier. De plus, CMÉ, tel qu'il est utilisé dans ce projet, fournit un processus de recherche itératif en quatre étapes basé sur sept principes, qui est dérivé de ADR et DSR ([101, 102]). Les étapes et les principes de l'approche ADR sont présentés plus en détail ci-dessous.

3.1.1 Étape 1 : Formulation du problème

L'ADR commence par identifier un problème au sein d'une organisation. Ce problème peut provenir de différentes parties prenantes, découler de recherches empiriques ou reposer sur des besoins identifiés sur le terrain [101]. Au cours de l'étape de formulation du problème, les chercheurs et le partenaire industriel déterminent la portée, les rôles et les questions de recherche qui seront étudiés. Cela aide à identifier et à conceptualiser l'opportunité de recherche sur la base des connaissances existantes [101]. Dans ce projet, la formulation de la problématique a été menée en amont et a conduit à une première proposition de recherche afin de trouver des financements. Ensuite, le cahier des charges de recherche a été créé, qui est le présent chapitre. La formulation du problème a permis d'identifier les solutions possibles et les références de la littérature scientifique.

Il est important d'obtenir l'accord formel du partenaire industriel et de déterminer comment le projet de recherche s'inscrit dans la problématique globale de l'organisation afin de développer de nouvelles connaissances scientifiques. La collaboration entre les chercheurs et un partenaire industriel, soutenue par la méthodologie ADR, permet une compréhension mutuelle de la portée et des objectifs de la recherche. Ceci est lié au Principe 1 : « La recherche est basée sur la pratique » et au Principe 2 : « Les artefacts sont enracinés dans la théorie ».

Pour appliquer ces deux premiers principes de la méthodologie ADR, nous avons d'abord cherché à obtenir l'avis des professionnels de la conformité informatique et d'autres parties prenantes. Cela s'est traduit par des entretiens, des réunions de réflexion et des groupes de discussion afin de mieux comprendre les défis et les besoins de la gestion de la conformité dans la pratique. Cela garantit que notre recherche est ancrée dans les réalités du secteur et que l'outil d'aide à la décision proposé aborde des problèmes concrets.

Outre cette approche basée sur la pratique, il est également essentiel de fonder notre artefact de recherche sur la théorie. À cette fin, nous nous sommes appuyés sur des théories et des cadres pertinents tirés de la littérature, tels que le modèle d'intégration de la maturité des capacités (Capability Maturity Model Integration (CMMI)) [103] et le cadre des objectifs de contrôle pour les technologies de l'information et connexes (Control Objectives for Information and Related Technologies (COBIT)) [104]. Ces théories ont contribué à la conception et au développement de l'outil d'aide à la décision, en veillant à ce qu'il soit basé sur les meilleures pratiques et les principes établis de la gestion de la conformité des technologies de l'information. Le Tableau 3.1 représente les activités de l'étape en détail.

TABLEAU 3.1 Activités de l'Étape 1 : Formulation du problème.

Ce tableau clarifie les activités réalisées durant l'étape 1 : Formulation du problème. Les activités 1.1-1.8 sont reliées au Principe 2 : « Les artefacts sont enracinés dans la théorie ». Les activités 1.9-1.12 sont reliées au Principe 1 : « La recherche est basée sur la pratique ».

Réf	Activité(s)
1.1	Début du projet
1.2	Analyse du marché actuel
1.3	Analyse des besoins du projet
1.4	Comparaison des produits et étalonnage des performances
1.5	Détermination des choix techniques à développer
1.6	Proposition de recherche initiale
1.7	Identification des composantes du problème
1.8	Revue de littérature
1.9	Suggestions de solutions possibles
1.10	Discussions initiales sur le projet
1.11	Spécifications de recherche et document de conception
1.12	Approbation du cahier des charges de la recherche et du document de conception

3.1.2 Étape 2 : Construction, Mise en œuvre et Évaluation (CMÉ)

Dans l'Étape 2, nous visons à développer des principes de conception généralisables en évaluant le problème et l'artefact, ce qui permet de clarifier progressivement l'espace d'innovation ciblé, pouvant être issu de la conception de l'artefact ou de l'intervention. Pour ce faire, nous utilisons la problématique et les fondements théoriques développés lors de l'Étape 1 comme point de départ pour la conception de l'artefact. Nous évaluons ensuite à la fois le problème et l'artefact pour identifier les principaux artefacts construits, qui sont représentés par les rectangles en bleu dans la Figure 3.1. Ces artefacts comprennent la revue de littérature identifiant les lacunes dans le domaine de conformité, l'ontologie développée à partir de nos recherches sur les cadres de conformité et de gouvernance, ainsi que le prototype final de la solution. Nous avons également identifié des artefacts intermédiaires, représentés par les rectangles en vert dans la Figure 3.1, tels que les métriques identifiées dans la littérature, les journaux et registres informatiques, les outils ontologiques construits à partir de l'ontologie, ainsi que le modèle de classification utilisé pour catégoriser et labéliser les données selon leur statut de conformité. En suivant les principales itérations représentées par les cercles jaunes dans la figure, nous avons développé un inventaire des métriques de conformité, quantifié et bonifié ces métriques, intégré les composantes de la solution et développé et validé le prototype.

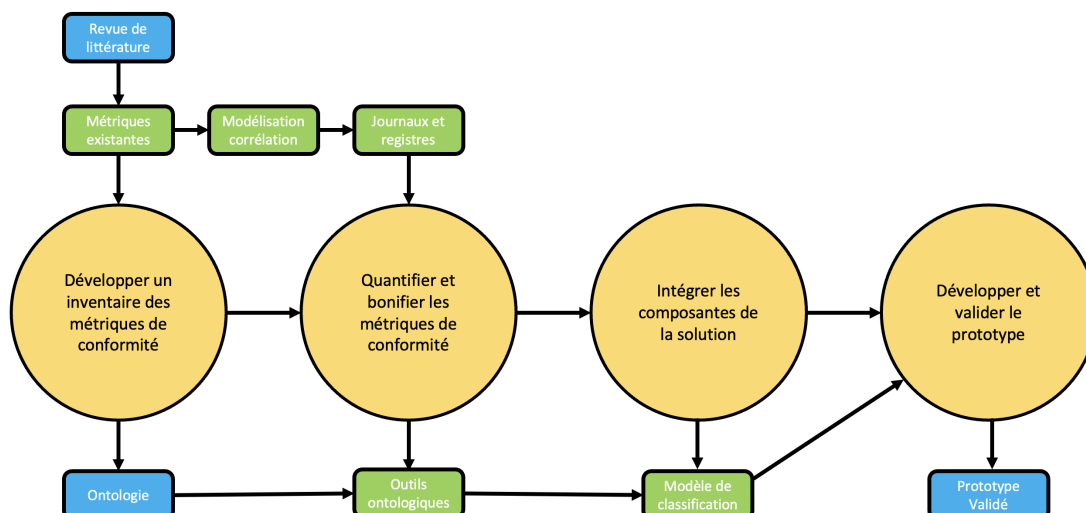


FIGURE 3.1 Description du cycle CMÉ lié à ce projet.

Les rectangles en bleu représentent les principaux artefacts construits. Les rectangles en vert correspondent aux artefacts intermédiaires. Les cercles jaunes correspondent aux principales itérations suivies.

Le Tableau 3.2 détaille les différentes activités de l'Étape 2. Le cycle CMÉ est un processus itératif qui aide à créer une solution technologique innovante dès le départ, contribuant ainsi à la mise en œuvre de la méthodologie. L'artefact et les théories émergentes qui sont intégrés dans le cycle de CMÉ seront constamment instanciés et testés à plusieurs reprises grâce à une intervention organisationnelle, et soumis aux hypothèses, aux attentes et aux connaissances des chercheurs. Ce processus renforce l'engagement organisationnel et guide la conception finale de l'artefact. Sur la base de ces interactions initiales, l'équipe de recherche intègre un artefact plus mature dans un cadre organisationnel plus large afin de permettre son évaluation en cours d'utilisation. Cela permet une amélioration continue de l'artefact à mesure qu'il est façonné et remodelé par le contexte d'utilisation. Dans ce projet de recherche, le cycle de CMÉ s'est terminé lorsqu'un prototype fonctionnel du système automatisé d'évaluation de la conformité a été terminé. Au cours de la phase d'évaluation du cycle CMÉ, la validité et l'utilité de l'outil sont évaluées [101]. Cette deuxième étape de l'ADR est basée sur les principes suivants : Principe 3 : "Mise en forme réciproque", Principe 4 : "Rôles d'influence mutuelle" et Principe 5 : "Évaluation authentique et simultanée".

En particulier, le Principe 3 implique que l'artefact et le contexte d'utilisation s'influencent mutuellement et sont co-constitués au fil du temps. Cela signifie que notre solution technologique est constamment ajustée en fonction de l'évolution des besoins organisationnels et des connaissances acquises lors de son utilisation. Le Principe 4 souligne l'importance de la col-

laboration entre les parties prenantes dans le développement et l'utilisation de l'artefact. La co-cr  ation est essentielle pour s'assurer que l'outil r  pond aux besoins de toutes les parties prenantes impliqu  es dans le processus de conformit  . Le Principe 5 met en avant l'importance de l'  valuation continue de l'artefact en cours d'utilisation, de mani  re transparente et ouverte, pour s'assurer qu'il reste efficace et pertinent.

TABLEAU 3.2 Les activit  s de l'  tape 2 : Construction, Mise en   uvre et   valuation (CM  ). Ce tableau clarifie les activit  s r  alis  es durant l'  tape 2. Les activit  s 2.1 et 2.7 sont reli  es au Principe 3 : "Mise en forme r  ciproque". Les activit  s 2.4 et 2.5 sont reli  es au Principe 4 : "R  les d'influence mutuelle". Les activit  s 2.2, 2.3, 2.6, 2.8, 2.9 et 2.10 sont reli  es au Principe 5 : "  valuation authentique et simultan  e".

R��f	Activit��(s)
2.1	Identification, documentation et validation des m��triques de conformit�� existantes
2.2	Analyse des journaux et des alertes
2.3	Identification des sources de donn��es pour les m��triques dans les journaux
2.4	Mod��lisation, agr��gation, corr��lation puis une validation par simulation des m��triques
2.5	Discussions et remue-m��ninges avec le partenaire industriel et l'��quipe de recherche
2.6	Analyse approfondie des ��l��ments identifi��s dans la revue de la litt��rature
2.7	Identification des ��l��ments de conformit�� �� int��grer dans la solution
2.8	Analyse des alertes de conformit�� fournies par l'outil d��velopp��
2.9	Gestion des non-conformit��s
2.10	D��veloppement et validation d'une preuve de concept

3.1.3   tape 3 : R  flexion et apprentissage

Lors de l'  tape 3 de l'ADR, l'  quipe examine si la solution propos  e pour un cas particulier peut   tre g  n  ralis  e pour   tre appliqu  e    un probl  me plus vaste. La d  marche de recherche vise    aller au-del   de la simple r  solution d'un probl  me ponctuel en utilisant la r  flexion sur le probl  me, les th  ories et les connaissances   mergentes pour identifier les contributions    l'avancement des connaissances. L'  tape 3 est bas  e sur le Principe 6 de l'ADR, "Emergence guid  e", qui vise    maintenir une orientation claire tout en permettant l'  mergence de nouvelles id  es. Dans notre recherche, nous avons utilis   ce principe pour rester align   sur notre objectif initial tout en   tant suffisamment flexible pour permettre l'  mergence de nouvelles id  es. Nous avons adapt   notre m  thodologie en fonction de nos r  sultats ant  rieurs et des nouvelles connaissances acquises pour guider l'  mergence de nouvelles id  es et rester pertinents pour la communaut   des professionnels de la conformit   TI. Le Tableau 3.3 d  taille les diff  rentes activit  s de l'  tape 3.

TABLEAU 3.3 Les activités de l'Étape 3 : Réflexion et apprentissage.

Ce tableau clarifie les activités réalisées durant l'Étape 3. Les activités 3.1-3.8 sont toutes reliées au Principe 6 : "Emergence guidée".

Réf	Activité(s)
3.1	Identification de la solution proposée
3.2	Revue des solutions
3.3	Test Alpha
3.4	Réflexion sur le problème
3.5	Identification des contributions
3.6	Enrichir les modèles de conformité
3.7	Sélection des techniques utiles pour le projet
3.8	Adaptation du processus de recherche

3.1.4 Étape 4 : Formalisation de l'apprentissage

La quatrième étape de l'ADR consiste à formaliser l'apprentissage, c'est-à-dire à traduire les enseignements tirés des projets de recherche en solutions applicables à des problèmes similaires dans d'autres organisations. Ces résultats sont qualifiés de principes de conception et de théories raffinées, conformément au Principe 7 : "Généralisation des résultats".

Le Principe 7 vise à généraliser les résultats obtenus pour avoir un impact plus large sur la communauté de recherche et de pratique. Dans notre recherche, nous avons utilisé ce principe pour décrire les principes de conception de notre outil de gestion de la conformité TI et élaborer une théorie raffinée basée sur notre expérience dans le développement de l'outil. Cette théorie a été utilisée pour guider la conception de futurs outils de gestion de la conformité TI dans d'autres organisations, tout en prenant en compte leur contexte unique. Le Tableau 3.4 détaille les différentes activités de l'Étape 4.

TABLEAU 3.4 Les activités de l'Étape 4 : Formalisation de l'apprentissage.

Ce tableau clarifie les activités réalisées durant l'Étape 4. Les activités 4.1-4.5 sont toutes reliées au Principe 7 "Généralisation des résultats".

Réf	Activité(s)
4.1	Développer des leçons
4.2	Évaluation de l'avenir des techniques à développer
4.3	Itération pour l'amélioration et Analyse SWOT
4.4	Rédaction du rapport final
4.5	Clôture - Rédaction du mémoire et transfert de technologie

3.2 Architecture de la solution

3.2.1 Approche méthodologique pour l'architecture de la solution

L'étape précédente de notre recherche a été consacrée à la collecte de données à travers des entretiens semi-structurés et l'analyse de la littérature pour comprendre les besoins en matière de gestion de la conformité TI et les défis auxquels les organisations sont confrontées. Cette étape nous a permis de définir les exigences de notre solution et d'élaborer une conception initiale. Cependant, la description de la méthodologie de recherche (Cf. Section 3.1) ne fournit pas suffisamment d'informations sur la manière dont nous avons appliqué cette méthodologie pour parvenir à la solution finale. Dans cette section, nous allons expliquer en détail comment nous avons utilisé les enseignements tirés de l'étape précédente pour développer l'architecture de notre solution de gestion de la conformité TI, en utilisant les principes de conception et les théories raffinées résultant de l'application des Principes 1-7 de l'ADR. En particulier, les composants suivants ont été identifiés pour inclusion dans notre solution :

- un système de gestion des règles et des contrôles de conformité informatique,
- un processus automatisé d'audit de conformité à partir des journaux de système et des éléments de l'infrastructure de sécurité,
- un entrepôt de données de conformité,
- un processus d'analyse de l'état des configurations par rapport aux exigences de conformité,
- un processus automatisé d'analyse de l'état de la conformité, et
- un tableau de bord de la conformité pour les gestionnaires.

3.2.2 Spécificités de l'architecture du projet

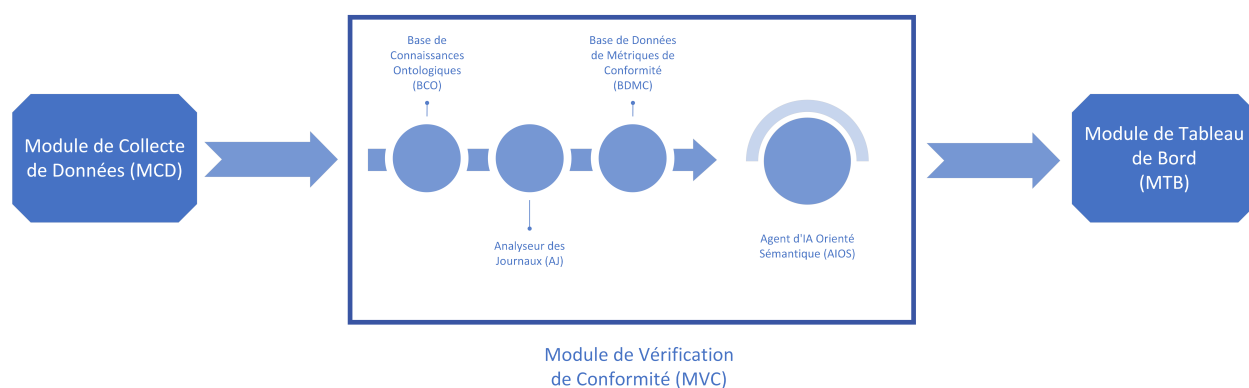


FIGURE 3.2 Architecture globale du projet

La Figure 3.2 présente l'architecture globale de notre solution, qui est composée des éléments suivants :

1. **Le Module de Collecte de Données (MCD) :** La collecte de données est la première étape cruciale de notre solution de gestion de la conformité TI. Elle nécessite une approche efficace et efficiente pour s'assurer que toutes les données nécessaires sont collectées et que les ressources sont utilisées de manière optimale. Pour répondre à ces besoins, nous avons développé le MCD qui est responsable de la collecte des données, de leur augmentation si nécessaire et de leur préparation avant qu'elles ne soient utilisées pour l'analyse ou pour nourrir le MVC. Le module utilise des méthodes de collecte de données bien établies comme décrit dans la section 3.3.
2. **Le Module de Vérification de Conformité (MVC) :** Ce module est au cœur de la solution et constitue la partie la plus importante. Il est doté de toutes les connaissances nécessaires sur les contrôles issus des normes ainsi que de l'IA nécessaire pour la prise de décision. Le MVC utilise cette intelligence et les informations collectées par le MCD pour déterminer si le système audité est conforme ou non. Il se compose de quatre sous-parties :
 - la Base de Connaissances Ontologiques (BCO),
 - l'Analyseur des Journaux (AJ),
 - la Base de Données de Métriques de Conformité (BDMC),
 - et l'Agent de l'Intelligence artificielle Orienté Sémantique (AIOS).

La BCO est un référentiel de connaissances qui contient des définitions formelles des concepts et des relations entre eux pour représenter les connaissances sur les normes et les réglementations. L'AJ traite les journaux de système collectés pour extraire les événements pertinents pour la conformité. La BDMC stocke des métriques de conformité pertinentes pour permettre une comparaison entre les résultats obtenus. Enfin, l'AIOS est un sous-module d'apprentissage automatique basé sur une ontologie interprétable qui permet de modéliser les connaissances sur la conformité et d'effectuer des vérifications de conformité en utilisant ces connaissances. Les sections suivantes fourniront des détails sur chacun de ces sous-modules : la section 4.1 pour la BCO, la section 4.2 pour l'AJ, la section 4.3 pour la BDMC, et la section 5.2 pour l'AIOS.
3. **Le Module de Tableaux de Bord (MTB) :** est une interface utilisateur qui permet d'afficher les informations de conformité de différentes manières. Notre solution a été conçue avec un tableau de bord Web simple et convivial qui permet aux utilisateurs d'obtenir facilement un aperçu de l'état de conformité de leur système. Le MTB permet de générer un rapport Excel contenant des informations sur la conformité, qui peut être utilisé pour une analyse plus approfondie de la conformité. Le MTB

prend ces informations et les affiche sous forme de graphiques et de tableaux faciles à comprendre, qui peuvent être utilisés pour évaluer rapidement l'état de conformité du système. Les détails sur le fonctionnement et la conception de ce module sont expliqués dans la section 5.6.

3.3 Module de Collecte de Données (MCD)

3.3.1 Élaboration du Module de Collecte de Données

Afin de développer le MCD adapté à notre système et répondre à nos questions de recherche, nous avons procédé selon les étapes suivantes :

1. Sélection des sources de données pertinentes, incluant notre partenaire industriel, des normes publiques (ISO 27002 :2013 [9], PCI-DSS v3.2.1 [10] et NIST-SP 800-53 rev-5 [11]), des ontologies de cybersécurité, le jeu de règles Wazuh [105–107], et un jeu de données personnalisé basé sur la commande Linux logger [108, 109].
2. Choix des méthodes de collecte, combinant approches manuelles et automatisées via scripts et outils.
3. Conception d'un plan de collecte détaillé, précisant les modalités et la fréquence de collecte pour chaque source.
4. Mise en œuvre du MCD en configurant les outils et scripts nécessaires conformément au plan de collecte.
5. Surveillance et maintenance du MCD pour garantir son bon fonctionnement, ainsi que la fiabilité et l'exactitude des données collectées, en effectuant des mises à jour et des modifications si nécessaire.
6. Validation des données pour assurer leur exactitude et exhaustivité à travers un processus incluant la comparaison avec d'autres sources, la détection d'erreurs, de valeurs aberrantes et le contrôle qualité.
7. Analyse des données pour identifier des tendances, des schémas et des idées susceptibles de répondre à nos questions de recherche.

3.3.2 Contrôles de sécurité sélectionnés

Nous avons sélectionné un type de contrôles spécifiques pour chacune des trois normes ISO 27002 :2013 [9, 110], PCI-DSS v3.2.1 [10] et NIST-SP 800-53 rev-5 [11] :

- PCI-DSS : "PCI_Requirement_1.1", établissant et mettant en œuvre des normes de configuration de pare-feu et de routeur pour contrôler les entrées et les sorties réseau.

- NIST-SP-800-53 : "AC-2 Account Management", définissant et documentant les types de compte, les gestionnaires de compte, les conditions préalables, les utilisateurs autorisés, les autorisations d'accès, et la création, activation, modification, désactivation et suppression des comptes. Les processus de gestion des comptes doivent être alignés sur les processus de résiliation et de transfert.
- ISO27002 : "Category 9 : Access Control", garantissant que seuls les utilisateurs autorisés ont accès à un service et empêchant les accès non autorisés. Cela requiert une politique et des processus de contrôle définis, l'enregistrement, la suppression et l'examen des droits d'accès, l'accès physique et au réseau, le contrôle des utilitaires privilégiés et la limitation de l'accès au code source du programme.

3.3.3 Données fournies par le partenaire industriel

La protection de la confidentialité des données est primordiale, et pour cela, nous avons assuré la dépersonnalisation et l'anonymisation des données fournies par notre partenaire industriel, soumises à une clause de confidentialité. Le processus rigoureux mis en place supprime toutes les informations d'identification, telles que noms, adresses et numéros de téléphone.

Bien que les détails spécifiques ne puissent être divulgués, les données reçues concernent le contrôle d'accès, l'authentification et la gestion des utilisateurs. Ces informations essentielles à l'analyse de la posture de sécurité ont été déterminantes pour développer et tester notre Module de Collecte de Données (MCD), assurant ainsi l'exactitude et la fiabilité des données collectées.

Les données incluaient également des scans de sécurité effectués avec les outils CloudSploit, ScoutSuite et Prowler [111–113]. Ces scanners de sécurité et de configuration sont conçus pour détecter les problèmes de sécurité potentiels dans les environnements infonuagiques, principalement sur l'infrastructure Amazon Web Services (AWS). Ils permettent d'identifier les risques de sécurité et d'aider les utilisateurs à comprendre comment les atténuer.

3.3.4 Génération de journaux avec Logger

Les journaux de conformité sont essentiels pour analyser des éléments tels que le contrôle d'accès, l'activité des utilisateurs et les configurations du système, en lien avec les contrôles sélectionnés dans 3.3.2. La création de journaux de conformité personnalisés permet de capturer et d'étiqueter les données relatives aux cas de conformité et de non-conformité, qui

peuvent servir de preuves lors d’audits ou d’enquêtes. Nous avons utilisé la commande `Logger` [108, 109] pour générer des données étiquetées destinées à notre pipeline d’AA.

Les journaux générés couvrent divers événements et scénarios, tels que les enregistrements d’authentification des utilisateurs, le type de connexion établie (par exemple, authentification multifactorielle par SMS ou clé de sécurité physique), et le statut de la connexion. Ils incluent également des enregistrements concernant la configuration de points d’accès, d’unités de stockage et de grappes de calcul, ainsi que leur état de chiffrement et leur conformité à divers critères. Les activités des utilisateurs, comme la définition des mots de passe, l’accès aux systèmes et ressources humaines, les tentatives d’authentification échouées, les verrouillages de compte et les déconnexions pour cause d’inactivité, sont également documentées.

De plus, les journaux traitent de la configuration des composants réseau tels que les pare-feu, les routeurs et les commutateurs, de la sauvegarde des fichiers d’audit sur un serveur de journalisation centralisé, ainsi que de la révocation des accès privilégiés, de la désactivation des comptes, de la restriction des référentiels de données, de l’isolation des fonctions de sécurité et de la programmation des sauvegardes.

3.3.5 Adaptation des règles Wazuh pour la conformité

Wazuh est une plateforme de sécurité open-source [105–107] conçue pour surveiller et gérer les alertes de sécurité générées par divers systèmes et applications au sein d’un réseau. Les règles Wazuh offrent un ensemble de règles prédéfinies pour se conformer aux exigences réglementaires de notre projet. Elles sont hautement personnalisables et permettent d’enrichir et d’améliorer nos journaux de conformité créés avec `Logger`.

L’ensemble de règles Wazuh contient des scénarios et des schémas prédéfinis pour les journaux, facilement liés aux contrôles choisis dans 3.3.2. Les règles peuvent être adaptées pour détecter les anomalies, les violations de politique et les activités suspectes, offrant ainsi une couverture complète des aspects de sécurité et de conformité. Les règles sont également mises à jour régulièrement pour s’adapter aux nouvelles menaces et aux changements réglementaires.

La personnalisation des règles Wazuh a permis de créer des alertes spécifiques à notre projet, en fonction des normes et des contrôles de sécurité sélectionnés. Par exemple, nous avons créé des alertes pour détecter les accès non autorisés, les modifications de configuration

non conformes et les tentatives de violation des politiques d'accès. En outre, la plateforme Wazuh facilite l'intégration avec d'autres outils de surveillance et de gestion de la sécurité, tels que les systèmes de détection d'intrusion et les outils d'analyse des journaux.

3.3.6 Utilisation d'ontologies open-source pour la conformité

Le Module de Collecte de Données (MCD) emploie plusieurs ontologies open-source liées à la conformité pour rassembler, analyser et traiter les données correspondant aux contrôles sélectionnés dans 3.3.2. Ces ontologies sont essentielles pour la mise en conformité globale, offrant un moyen normalisé de représenter et d'interpréter les données de conformité. Pour assurer une gestion et un partage efficaces des données, nous avons respecté les principes FAIR (Findable, Accessible, Interoperable, and Reusable) [114] lors de la collecte des ontologies et de la construction du MCD.

De plus, nous avons adopté la méthodologie "Ontology-Based Access Control (OBAC)" [115] en combinaison avec les principes FAIR pour les raisons suivantes :

1. L'OBAC permet un accès granulaire aux données en utilisant des protocoles basés sur l'ontologie et des métadonnées. Cette approche garantit que les utilisateurs n'accèdent qu'aux informations pour lesquelles ils sont autorisés, en tirant parti des technologies du web sémantique pour définir les politiques de sécurité.
2. Les politiques OBAC sont réutilisables et facilitent l'intégration de plusieurs systèmes d'information en évitant de se focaliser sur les détails d'implémentation. Cette abstraction rend la spécification et la gestion des politiques de sécurité plus systématiques et faciles à gérer. L'OBAC autorise également des politiques qui permettent l'accès à des données conformes à des modèles spécifiques dans l'ontologie mise en œuvre, avec des données sous-jacentes représentées sous forme de triples RDF dans un catalogue varié.

Dans les sections suivantes, nous présentons les cinq ontologies/modèles ontologiques utilisés dans notre approche.

Adaptation et configuration du modèle ontologique OBACM

Le premier modèle ontologique que nous avons adapté à notre solution est l'Ontology-Based Access Control Model (OBACM) proposée par [116]. L'OBACM étend le contrôle d'accès basé sur les attributs et les tâches sensibles au contexte (Context Sensitive Attribute and Task-Role Based Access Control (CSAT-RBAC)) [116] pour les applications de services web et

simplifie la gestion de la sécurité pour les utilisateurs dynamiques et anonymes. Ce modèle de contrôle d'accès sémantiquement compatible fournit des représentations ontologiques pour les concepts et les relations impliqués dans le modèle CSAT-RBAC, avec chaque composant représenté dans une ontologie distincte.

Nous avons apporté des modifications et des améliorations à l'ontologie OBACM pour mieux l'adapter à notre solution de conformité. En particulier, nous avons veillé à ce que le modèle applique strictement le principe du moindre privilège dans l'attribution des rôles et n'accorde des privilèges qu'aux utilisateurs justifiés. Nous avons également optimisé les instances de services web et les instances de tâches pour permettre un contrôle d'accès précis et granulaire.

Pour prendre en compte l'hétérogénéité et la dynamique des utilisateurs du web, nous avons intégré le contrôle d'accès basé sur les justificatifs et l'attribution dynamique des rôles. Ces ajustements ont permis de sécuriser le travail administratif manuel en spécifiant les autorisations pour chaque sujet de sécurité par rapport à chaque objet de sécurité, simplifiant ainsi la gestion de la sécurité et la rendant plus efficace.

Enfin, nous avons adapté l'ontologie de processus pour représenter l'attribution des rôles d'utilisateur et l'ontologie de politique pour contrôler l'accès à l'attribution des rôles d'utilisateur. Ces ajustements ont renforcé l'efficacité de notre solution de conformité en utilisant OBACM.

Adaptation et configuration de l'ontologie BACLO

La deuxième ontologie que nous avons intégrée et adaptée à notre solution est l'ontologie de contrôle d'accès de base (Basic Access Control Ontology (BACLO)) [117, 118]. BACLO est basée sur le système distribué WebAccessControl [117, 118] et conçue pour contrôler l'accès aux ressources en ligne en se basant sur les URI HTTP. Elle permet de gérer les autorisations pour les utilisateurs et les groupes avec précision et facilité.

Pour adapter BACLO à notre solution, nous avons d'abord identifié les termes nécessaires au stockage des listes de contrôle d'accès. En utilisant cette ontologie comme base, nous avons réduit la redondance dans la mise en œuvre des mécanismes de contrôle d'accès et facilité le partage du code entre les différents modules de notre système.

Nous avons ensuite modifié l'ontologie pour faciliter la définition, le stockage et l'application des politiques de contrôle d'accès dans notre solution. Grâce à ces ajustements, nous avons pu renforcer la sécurité des ressources en ligne et contrôler l'accès aux données et aux ressources de manière plus granulaire.

L'intégration de l'ontologie BACLO nous a permis de standardiser la représentation des politiques de contrôle d'accès et d'assurer leur compatibilité avec diverses applications et domaines.

Adaptation et configuration de l'ontologie AMO

La troisième ontologie que nous avons adaptée à notre solution est l'Access Management Ontology (AMO) [119]. AMO est conçue pour représenter les droits d'accès accordés à certains utilisateurs ou classes d'utilisateurs sur un document. Elle repose sur des principes communs aux systèmes de gestion de contenu, tels que les rôles, les actions autorisées et les types d'accès aux ressources.

Pour personnaliser l'AMO à notre solution, nous avons d'abord analysé les classes et propriétés essentielles de l'ontologie, y compris les classes de rôle, d'action et de type d'accès. Nous avons également examiné le vocabulaire Friend-Of-A-Friend (FOAF) [120] pour les classes d'agent, de groupe et de document, qui sont au cœur d'AMO.

Ensuite, nous avons ajusté l'ontologie pour qu'elle corresponde à nos besoins en matière de contrôle d'accès. Cela inclut l'ajout et la modification de classes et de propriétés pour mieux refléter notre solution de conformité et ses exigences. En intégrant ces modifications, nous avons créé une base solide pour la gestion des droits d'accès au sein de notre système.

Enfin, nous avons incorporé des règles d'inférence spécifiques à notre contexte pour représenter de manière déclarative notre stratégie de contrôle d'accès. Cette personnalisation a rendu notre gestion des droits d'accès facilement adaptable aux changements de politique sans affecter les annotations des ressources contrôlées.

En adaptant l'ontologie AMO à notre solution, nous avons pu intégrer un système de contrôle d'accès simple et efficace pour gérer les droits d'accès aux documents et aux ressources de notre système de conformité.

Adaptation et configuration d’une ontologie basée sur PCI-DSS

La quatrième ontologie que nous avons adaptée provient de [121] et représente les règles mandatées par Payment Card Industry- Data Security Standard (PCI-DSS). Cette ontologie offre une représentation riche et traitable par machine, permettant de conceptualiser, créer et valider les contrôles de conformité PCI-DSS.

Pour intégrer cette ontologie à notre solution, nous avons examiné le graphe de connaissances associé et identifié les obligations pertinentes pour notre contexte. Nous avons également étudié les liens entre les contrôles PCI-DSS et ceux de la Cloud Security Alliance (CSA) [21].

Nous avons ensuite personnalisé l’ontologie pour l’adapter à nos besoins spécifiques, en ajoutant ou modifiant des éléments si nécessaire. Cette adaptation nous a permis de bénéficier d’une ontologie précise et exhaustive pour notre solution de conformité.

Adaptation et configuration d’une ontologie basée sur ISO 27002

La dernière ontologie que nous avons incorporée provient de [58] et formalise la norme ISO/IEC 27002 :2013. Cette ontologie fournit une base de connaissances pour automatiser les projets de conformité en matière de sécurité de l’information.

Afin d’adapter cette ontologie à notre solution, nous avons analysé la méthode proposée pour cartographier les directives de sécurité informelles en descriptions formelles. Nous avons également étudié les règles formelles de mise en œuvre des contrôles développées par les auteurs.

Par la suite, nous avons ajusté l’ontologie pour qu’elle corresponde à nos exigences et notre contexte. Nous avons ajouté ou modifié des éléments de l’ontologie en fonction des besoins spécifiques de notre solution de conformité.

CHAPITRE 4 MODULE DE VÉRIFICATION DE CONFORMITÉ (MVC)

Le Module de Vérification de Conformité (MVC) (Cf. Figure 4.1) représente un élément essentiel de notre solution intégrée conçue pour aider les organisations à gérer efficacement leur conformité des TI. Dans ce chapitre, nous présentons le MVC en détaillant l'enchaînement de ses composants, ses objectifs, ses principales fonctionnalités et son importance au sein de la solution globale. De plus, nous aborderons les défis spécifiques auxquels le MVC cherche à répondre et les avantages qu'il offre pour optimiser la gestion de la conformité des TI.

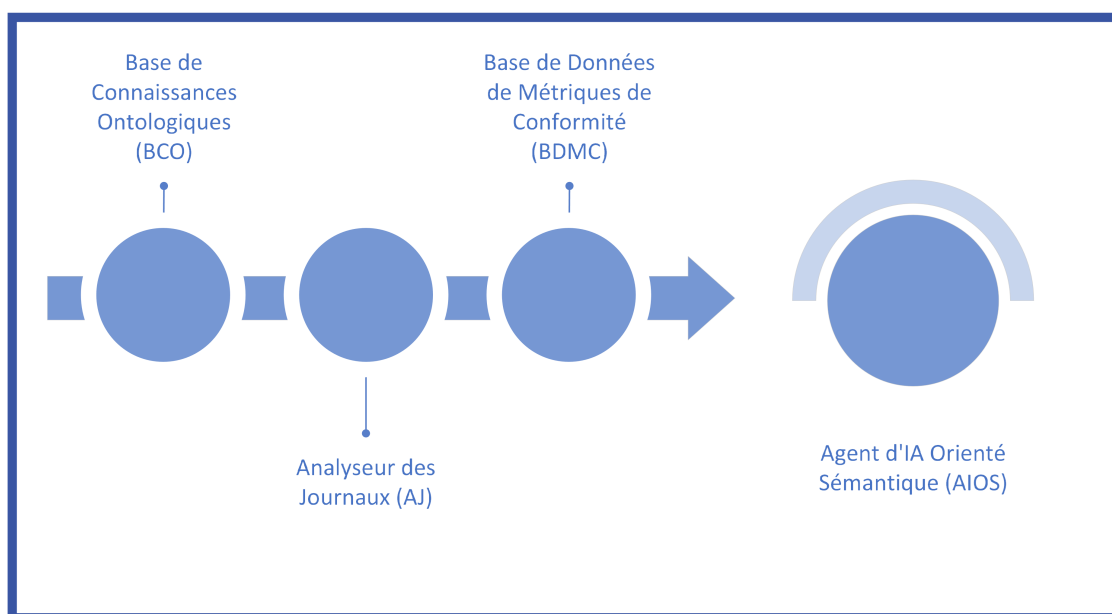


FIGURE 4.1 Composants du Module de Vérification de Conformité (MVC).

MVC est composé de la Base de Connaissances Ontologiques (BCO), de l'Analyseur des Journaux (AJ), de la Base de Données de Métriques de Conformité (BDMC) et de l'Agent de l'Intelligence artificielle Orienté Sémantique (AIOS).

4.1 La Base de Connaissances Ontologiques (BCO)

Dans cette section, nous abordons le développement d'une ontologie conçue pour représenter les contrôles de conformité, en mettant l'accent sur les contrôles techniques et organisationnels. L'ontologie a été élaborée avec une flexibilité permettant l'intégration de futures réglementations et normes. Nous expliquons la méthodologie de création de l'ontologie, y

compris l'application de la théorie pratique de la modélisation des connaissances et les outils utilisés durant le développement. Le chapitre retrace les différentes étapes du processus, depuis la définition des objectifs et du périmètre de l'ontologie jusqu'à la collecte d'informations, le raisonnement et la formalisation de l'ontologie. Les exigences de l'ontologie ont été régulièrement réexaminées tout au long de sa conception, en étroite relation avec la gestion des besoins.

4.1.1 Objectifs et portée de l'ontologie

À cette étape, nous orientons le modèle de connaissances en définissant les caractéristiques essentielles de l'ontologie, telles que le domaine d'intérêt, l'objectif global, les principales parties prenantes et la portée. L'ontologie en développement se concentre sur le contrôle d'accès, les pare-feux et la gestion des comptes selon les normes ISO27002, PCI-DSS et NIST-SP-800-53. Bien que ciblant le domaine réglementaire, elle ne couvre pas tous les aspects de la conformité des TI. Toutefois, notre ontologie flexible et extensible sert de base pour des travaux futurs dans ce domaine. L'ontologie vise à déterminer les contrôles de sécurité nécessaires pour atteindre la conformité à une réglementation spécifique, en tenant compte des facteurs tels que la taille de l'entreprise et le type de données traitées. Les principales parties prenantes sont les consultants en conformité et les entreprises cherchant à se conformer aux réglementations. Les consultants jouent un rôle clé, car ils maintiendront le contenu de l'ontologie et s'y référeront.

4.1.2 Acquisition et structuration des connaissances

Cette étape implique la collecte et la sélection des connaissances pertinentes pour l'ontologie, en répertoriant les classes et propriétés initiales. Les méthodes de collecte d'informations dépendent de la nature des connaissances à modéliser. Dans notre cas, le brainstorming et l'analyse de contenu sont les plus appropriés. Nous analysons les sources de connaissances d'un point de vue sémantique, extrayons les classes et propriétés concrètes du texte brut non structuré et élaborons une carte mentale de l'ontologie préliminaire. Par ailleurs, nous avons fondé notre travail sur les ontologies repérées lors de la revue de littérature et au cours du processus de Module de Collecte de Données.

Lors de la structuration initiale, nous créons une liste des termes clés identifiés à partir des sources examinées. Cette liste permet de classer les termes en classes, instances, relations ou termes à écarter. Le résultat de cette phase est la classification de cette liste et la création d'une représentation visuelle de l'ontologie initiale sous forme de graphe. La liste complète

des termes utilisés dans ce travail d'ontologie est présentée en Annexe 2.

4.1.3 Formalisation et optimisation de l'ontologie

Notre ontologie a été formalisée en utilisant le langage OWL (Web Ontology Language) [47], un langage basé sur la syntaxe RDF (Resource Description Framework) [122]. Le RDF permet de formuler des déclarations logiques en termes de sujet-prédicat-objet, tandis que le langage OWL construit des relations sémantiques par-dessus. Au cœur de l'OWL se trouve le concept de "Thing", dont tous les autres concepts sont dérivés. Nous avons commencé par modéliser les classes, les relations de subsumption, les propriétés des objets et les propriétés des types de données de chaque classe. Ensuite, nous avons introduit des restrictions existentielles pour définir les conditions nécessaires de chaque classe [123]. Ces restrictions existentielles indiquent que si un individu appartient à une classe spécifique, il doit avoir certaines propriétés. Toutefois, la réciproque n'est pas vraie, c'est-à-dire qu'un individu ayant ces propriétés n'est pas nécessairement membre de cette classe. Cette caractéristique de l'OWL nécessite une attention particulière lors de l'édition du modèle de connaissances et peut permettre l'infiltration d'erreurs logiques dans l'ontologie. Pour surmonter ce défi, nous avons transformé autant de classes que possible en classes définies, en ajoutant des restrictions universelles sur toutes leurs propriétés d'objet [123]. Ces restrictions universelles définissent les conditions suffisantes pour qu'un individu appartienne à une classe. Enfin, nous avons appliqué des axiomes de fermeture à un maximum de classes, ce qui est considéré comme une bonne pratique [123]. L'axiome de fermeture, équivalent à une restriction universelle constituée de l'union de toutes les propriétés des objets d'une classe, permet au raisonneur de générer pour nous des relations de subsumption et de détecter des erreurs éventuelles dans la définition des relations d'héritage.

4.1.4 Structure détaillée de l'ontologie

Pour créer notre ontologie des frameworks de conformité informatique, nous avons adopté une approche hybride, en combinant l'utilisation d'ontologies préexistantes découvertes grâce au Module de Collecte de Données (MCD) (Cf. 3.3 et 3.3.6) et une méthode dynamique basée sur un fichier Excel structuré et un code Python.

Le fichier Excel, dont un aperçu est fourni dans la Figure 4.2, intègre un sous-ensemble des contrôles des normes de sécurité de l'information ISO 27002, NIST-SP 800-53 et PCI-DSS (Cf. 3.3.2) et comprend six colonnes caractérisant un contrôle de sécurité : notamment la classe à créer ("self"), sa classe parente ("parent"), l'étiquette ontologique ("label"), l'identifiant

	A	B	C	D	E	F
	self	parent	label	control_id	control_name	guidelines
1	PCI_Requirement_1.1.1	PCI_Requirement_1	PCI_Requirement_1.1.1	1.1	1.1	1.1 Establish and implement firewall and router configuration standards that include the following: 1.1 Inspect the firewall and router configuration standards and other documentation specified below and verify that standards are complete and implemented as follows: Firewalls and routers are key components of the architecture that controls entry to and exit from the network. These devices are software or hardware devices that block unwanted access and manage authorized access into and out of the network.
2	PCI_Requirement_1.1.1.1	PCI_Requirement_1	PCI_Requirement_1.1.1.1	1.1.1	1.1.1	1.1.1 A formal process for approving and testing all network connections and changes to the firewall and router configurations 1.1.1.a Examine documented procedures to verify there is a formal process for testing and approval of all: . Network connections and . Changes to firewall and router configurations 1.1.1.b For a sample of network connections, interview responsible personnel and examine records to verify that network connections were approved and tested. 1.1.1.c Identify a sample of actual changes made to firewall and router configurations, compare to the change records, and interview responsible personnel to verify the changes were approved and tested. A documented and implemented process for approving and testing all connections and changes to the firewalls and routers will help prevent security problems caused by misconfiguration of the network, router, or firewall.
3						

FIGURE 4.2 Aperçu du fichier Excel utilisé pour structurer les normes de sécurité de l'information, incluant les classes, les relations, les propriétés et les annotations.

unique ("control_id"), le nom ("control_name") et l'orientation ou les lignes directrices pour appliquer le contrôle ("guidelines"). Les informations contenues dans le fichier peuvent être connectées à d'autres ressources de sécurité de l'information grâce à l'utilisation de propriétés d'annotation et d'étiquettes ontologiques, ce qui permet d'obtenir une image plus complète et holistique du paysage de la sécurité de l'information. Notre code Python rationalise la procédure d'incorporation des composants du framework de conformité dans une ontologie OWL. Il supprime les doublons, génère de nouvelles classes et de nouveaux individus et effectue les inférences manquantes en convertissant le fichier Excel en un cadre de données Python (ou "Python Dataframe" en anglais). Des bibliothèques Python telles que Owlready2 [49] et RDFlib [124] et des outils comme l'éditeur Stanford Protégé ont facilité la création et la manipulation de l'ontologie. Nous avons ensuite généré la documentation de l'ontologie à l'aide de la commande python "ontospy" [125]. Notre ontologie offre une structure pour l'application des réglementations et facilite l'ajout de nouveaux frameworks. Complète, bien connectée et facilement représentée sous forme graphique, l'ontologie résultante est le fruit d'une recherche approfondie. L'automatisation des tâches de gestion de la conformité, telles que l'analyse des lacunes, la surveillance des problèmes et l'évaluation des risques, est possible grâce à elle.

L'approche hybride nous a permis de développer une ontologie sémantique qui fournit une structure complète pour modéliser la sécurité des actifs organisationnels. Nous avons inclus diverses classes et relations qui capturent les différents éléments du programme de sécurité

d'une organisation. L'ontologie comprend des classes telles que *Asset*, *Organization*, *Impact*, *SecurityAttribute*, *ThreatSource*, *ThreatOrigin*, *Vulnerability*, *SeverityScale*, *Control*, *ControlType*, *StandardControl*, *AccessType*, *Role*, *Action*, *UserCredential*, *UserCapability*, *UserID*, *Attribute*, *UserRoleAssignment*, *AccessControlStatus*, *Constraint* et *PermissionRoleAssignment*. Les relations entre ces classes sont bien définies. Par exemple, nous avons capturé les relations entre les actifs et les attributs de sécurité, les menaces et les impacts, les vulnérabilités et les contrôles, et les utilisateurs et leurs rôles et actions. L'ontologie est conçue pour aider les petites et moyennes organisations à développer leurs programmes de sécurité et peut être utilisée pour analyser progressivement les effets des contrôles de sécurité récemment mises en œuvre.

Concrètement, la classe, *Asset*, représente les actifs corporatifs et doit ainsi posséder un ou plusieurs attributs de sécurité pour se protéger des menaces potentielles. Les actifs peuvent également nécessiter d'autres actifs connexes ou en dépendance pour fonctionner correctement et sont la propriété d'une organisation. Des contrôles (i.e., les individus de la classe *Control*) peuvent être mis en œuvre sur les actifs pour atténuer les vulnérabilités potentielles, qui peuvent être exploitées par des menaces. Les menaces (i.e., les individus de la classe *Threat*) peuvent menacer les actifs et donner lieu à d'autres menaces et impacts. Les menaces peuvent avoir une source (i.e., la classe *Source*) et une origine (i.e., la classe *Origin*). La classe *Impact* affecte les actifs et les attributs de sécurité (i.e., la classe *Attribute*). Une échelle de gravité (i.e., la classe *SeverityScale*) est utilisée pour mesurer la gravité d'une vulnérabilité. Les types de contrôle (i.e., la classe *ControlType*) peuvent être définis comme préventifs, correctifs, dissuasifs, de récupération ou de détection, et correspondent aux contrôles standard. La classe *AccessType* peut être *Public*, *Private* ou *SemiPublic*, et est associé à un document qui a des actions autorisées pour différents rôles, y compris *Guest*, *Contributor* et *Administrator*. La classe *UserCredential* a des sous-classes *UserCapability* et *Attribute*. Les attributions *UserRoleAssignment* et *PermissionRoleAssignment* peuvent être utilisées pour attribuer diverses actions et statuts de contrôle d'accès aux utilisateurs en fonction de leur rôle et de leurs attributs. Tous les classes, individus et relations de l'ontologie se trouvent dans l'Annexe 7.4.

4.2 L'Analyseur des Journaux (AJ)

La méthodologie et l'architecture de la solution pour cette partie sont illustrées dans la Figure 4.3. L'approche a été adaptée et enrichie à partir des travaux de [88, 89, 94], qui proposaient initialement des modèles légers d'analyseurs de journaux pour détecter les activités malveillantes et les intrusions au système d'information. Dans notre étude, nous avons ap-

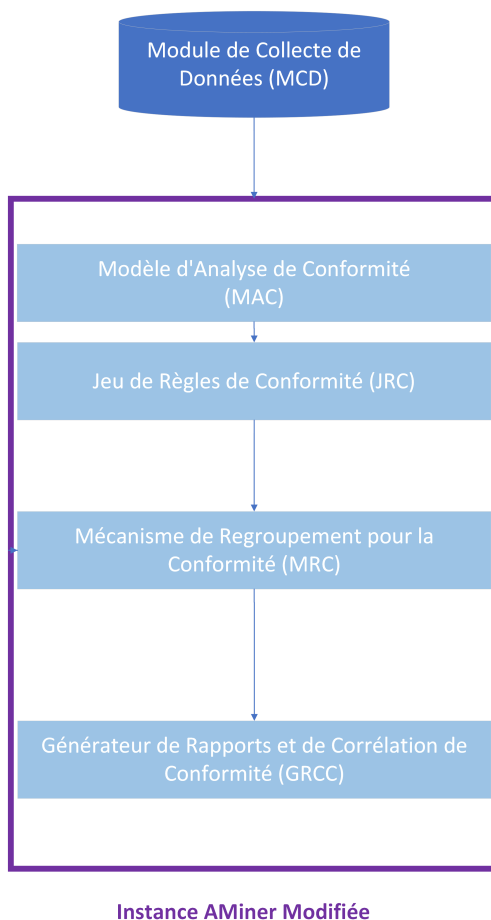


FIGURE 4.3 Pipeline de l'Analyseur des Journaux.

Ce schéma illustre le pipeline de l'Analyseur des Journaux, composé de quatre éléments principaux : le Modèle d'Analyse de Conformité (MAC), le Jeu de Règles de Conformité (JRC), le Mécanisme de Regroupement pour la Conformité (MRC) et le Générateur de Rapports et de Corrélations de Conformité (GRCC).

porté plusieurs contributions significatives et ajouté de la valeur aux travaux de [88, 89, 94], qui proposaient les outils Logdata Anomaly Miner (AMiner) et Automatic Event Correlation for Incident Detection (AECID). Inspirés par leur approche, nous avons étendu et enrichi leur méthode en ajoutant des règles et des composants algorithmiques adaptés au contexte de conformité. Nous avons réussi à entraîner et adapter les outils aux données de conformité, permettant une détection plus précise des violations de conformité aux normes dans les journaux. De plus, nous avons intégré la possibilité de générer des rapports de conformité bien structurés, facilitant l'analyse et la prise de décision. Enfin, nous avons ajouté de nouveaux hyperparamètres et métriques spécifiques au contexte de conformité, améliorant ainsi les performances de détection et élargissant le champ d'application de ces outils pour répondre aux

exigences des entreprises en matière de conformité aux normes et réglementations.

Le pipeline de l'Analyseur des Journaux (Cf. Figure 4.3) se compose de quatre éléments principaux : le Modèle d'Analyse de Conformité (MAC), le Jeu de Règles de Conformité (JRC), le Mécanisme de Regroupement pour la Conformité (MRC) et le Générateur de Rapports et de Corrélations de Conformité (GRCC). Ces quatre éléments représentent une instance AMiner [88, 89, 94] modifiée pour servir à notre projet de recherche.

Le Module de Collecte de Données (MCD) fournit les données des journaux. Le MAC est initialisé pour reconnaître les écarts indiquant une non-conformité, tels que les tentatives d'accès non autorisé, les champs de données manquants ou incorrects, ou les données ne respectant pas certaines exigences de formatage. Une fois formé, le MAC applique les inférences apprises aux nouvelles données de journaux pour détecter les cas potentiels de non-conformité, soit en temps réel, soit de manière périodique lors d'audits de conformité. L'aspect temps réel n'est pas couvert par notre étude.

Le pipeline interprète les journaux acquis en suivant le MAC spécifique et le JRC sur mesure, identifiant les événements considérés comme légitimes sur le système. Le MRC apprend le comportement normal (conforme) du système en analysant les journaux reçus. Le GRCC génère un rapport de décision concernant l'état de conformité des journaux reçus.

La configuration a été réalisée sur une machine virtuelle Ubuntu 20.04 LTS équipée de 8 cœurs de processeur, 32 Go de RAM et 250 Go de stockage. Certains tests ont également été effectués sur Google Colab, offrant gratuitement un GPU de 12 Go de RAM. L'implémentation originale [88, 89, 94], open source, est compatible avec Python 2 et 3 et a été testée sur CentOS 6.7 et Ubuntu Xenial 16.04. Dans sa configuration minimale, elle nécessite seulement 32 Mo de mémoire pour effectuer un simple filtrage des lignes de journaux. Si les exigences en matière de mémoire sont plus élevées, elle peut être exécutée sur une machine séparée, plus puissante, et écouter le flux de journaux provenant de l'appareil surveillé. Cela permet de traiter les journaux produits par des dispositifs à faible puissance de calcul, des systèmes anciens et des dispositifs dont le matériel ne répond pas aux exigences de l'installation.

Il convient de souligner que cette approche de détection de la non-conformité à travers les journaux doit être adaptée aux normes ou réglementations spécifiques appliquées, c'est-à-dire aux contrôles choisis décrits dans la section 3.3.2. Notre travail étend et améliore l'approche

originale en adaptant le pipeline pour détecter les violations de conformité, ajoutant ainsi de la valeur à la solution existante et élargissant son champ d'application.

4.2.1 Modèle d'Analyse de Conformité (MAC)

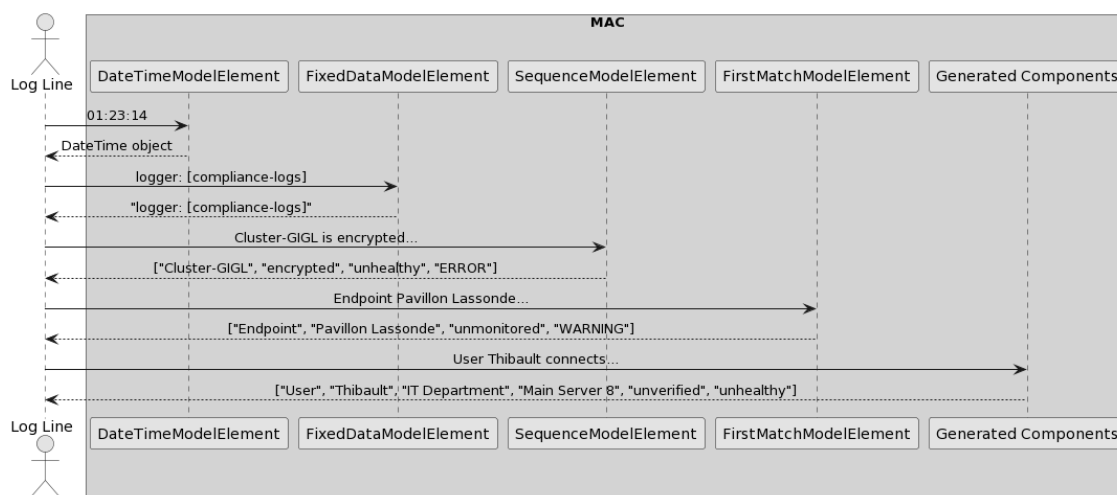


FIGURE 4.4 Exemple d'analyse d'une ligne de journaux par le Modèle d'Analyse de Conformité (MAC).

La figure illustre comment le MAC traite chaque ligne de journaux et applique des composants prédéfinis et générés pour extraire les informations pertinentes. Les lignes de journaux sont analysées à l'aide de composants prédéfinis : `DateTimeModelElement`, `FixedDataModelElement`, `SequenceModelElement`, `FirstMatchModelElement` et d'autres composants générés à dessein. Le résultat final est une interprétation des lignes de journaux, qui permet de détecter les problèmes de conformité.

Le pipeline utilise le Modèle d'Analyse de Conformité (MAC) pour interpréter les messages de journaux reçus du MCD. Le MAC, représenté sous forme de graphe orienté, permet de filtrer les informations redondantes avant d'effectuer une analyse approfondie. Les composants du MAC peuvent être prédéfinis ou spécifiquement générés (i.e., générés à dessein) pour l'analyse de conformité.

Les composants prédéfinis analysent divers types de données et correspondent à différentes structures dans les lignes de journaux. Ces composants prédéfinis comprennent :

- **AnyByteDataModelElement** : Correspond à tout ce qui se trouve jusqu'à la fin d'un atome-logarithme.
- **DateTimeModelElement** : Analyse l'heure de la date à l'aide du module `datetime` de Python

- **DecimalIntegerValueModelElement** (Élément de modèle de valeur décimale) : Analyse les valeurs entières
- **FirstMatchModelElement** (élément de modèle FirstMatch) : Branche le modèle en prenant la première branche correspondant aux données log-atomiques restantes
- **FixedDataModelElement** (Élément de modèle de données fixes) : Correspond à une chaîne de caractères fixe (constante)
- **IpAddressDataModelElement** (Élément de modèle d'adresse IP) : Correspond à une adresse IPv4
- **SequenceModelElement** (Élément de modèle de séquence) : Correspond à tous les sous-éléments exactement dans l'ordre donné
- **FixedWordlistDataModelElement** (Élément de modèle de données de liste de mots fixes) : Correspond à l'une des chaînes fixes d'une liste
- **VariableByteDataModelElement** (Élément de modèle de données à octets variables) : Correspond à des données de longueur variable encodées dans un alphabet donné
- **Base64StringModelElement** (Élément de modèle de chaîne de base64) : Analyse les chaînes base64 en tant que données binaires.
- **DateTimeModelElement** : Analyse simple des dates à l'aide du module python datetime. Voir aussi MultiLocaleDateTimeModelElement
- **DebugModelElement** : Affiche des informations de débogage lors de l'analyse d'un atome-logiciel.
- **DecimalFloatValueModelElement** : analyse des valeurs flottantes
- **DecimalIntegerValueModelElement** : analyse des valeurs intermédiaires
- **DelimitedDataModelElement** (élément de modèle de données délimitées) : Identique à l'élément AnyByteDataModelElement, mais les données ne sont incluses que jusqu'à la chaîne de délimitation donnée.
- **ElementValueBranchModelElement** : branchement conditionnel en raison des valeurs analysées au préalable.
- **HexStringModelElement** : Correspondance avec une chaîne hexadécimale.
- **MultiLocaleDateTimeModelElement** : Analyse les éléments temporels avec correction de l'année bissextile, prise en charge de plusieurs paramètres linguistiques.
- **OptionalMatchModelElement** (Élément de modèle à correspondance facultative) : Correspondance des sous-éléments zéro ou une seule fois.
- **RepeatedElementDataModelElement** (Élément de données répétées) : Fait correspondre les sous-éléments un nombre donné de fois
- **WhiteSpaceLimitedDataModelElement** (Élément de modèle de données limité

aux espaces) : Correspond à la chaîne de caractères jusqu'au prochain espace blanc.

Les composants générés à dessein sont créés en suivant ce processus : le MAC collecte un ensemble de lignes de journaux et applique soit un regroupement basé sur les mots, similaire à Simple Logfile Clustering Tool (SLCT) [126, 127], soit un regroupement basé sur les caractères, similaire à [128]. Ce regroupement permet d'identifier les parties statiques et variables des lignes de journaux. Les alignements obtenus, décrivant les groupes de lignes de journaux ainsi que les informations sur les segments de lignes de journaux apparaissant comme parties variables, sont utilisés pour construire le graphe décrivant le modèle de ligne de journaux.

Lorsque l'analyseur atteint un certain niveau de stabilité, c'est-à-dire lorsqu'il n'est plus nécessaire d'adapter le MAC après avoir analysé un nouvel ensemble de lignes de journaux, le pipeline met fin au processus de génération de composants. Enfin, les inférences apprises sont utilisées pour filtrer les lignes de journaux inconnues et générer des alertes de non-conformité. La Figure 4.4 présente un exemple d'analyse d'une ligne de journaux en détectant des problèmes de conformité avec le MAC. L'approche est illustrée sur les trois lignes de journaux suivantes :

- *01 :23 :14 logger : [compliance-logs] Cluster-GIGL is encrypted with status unhealthy : ERROR, encryption is broken or not properly established*
- *01 :23 :14 logger : [compliance-logs] Endpoint Pavillon Lassonde is started with status unmonitored : WARNING, endpoints should be monitored*
- *01 :23 :14 logger : [compliance-logs] User Thibault connects to IT Department Main Server 8 with access privileges unverified and status unhealthy*

Ainsi, le MAC offre une approche flexible et évolutive pour l'analyse de conformité des lignes de journaux. Il permet une adaptation dynamique à différentes structures de journaux et une détection efficace des problèmes de conformité. Cette approche ajoute de la valeur à l'analyse en améliorant la précision et la pertinence de l'interprétation des journaux, en réduisant les faux positifs et en mettant en évidence les problèmes potentiels de manière claire et structurée.

4.2.2 Jeu de Règles de Conformité (JRC)

Notre pipeline détecte les violations de conformité dans les messages de journaux en utilisant deux approches complémentaires : (i) l'identification des écarts par rapport au Modèle

d'Analyse de Conformité (MAC), et (ii) l'application d'un Jeu de Règles de Conformité (JRC) pour vérifier le respect des normes pertinentes.

La première approche repose sur la connaissance du comportement normal d'un système conforme, représenté par le Modèle d'Analyse de Conformité (MAC). En comparant les messages de journaux avec les chemins disponibles dans le graphe du MAC, le pipeline peut identifier les événements inattendus ou anormaux. Les messages de journaux conformes doivent correspondre à un chemin complet du graphe du MAC. Les écarts par rapport à ces chemins indiquent des événements de non-conformité, qui peuvent résulter de défaillances techniques, d'activités de maintenance ou d'attaques exploitant des fonctions inutilisées du système.

La seconde approche consiste à appliquer un Jeu de Règles de Conformité (JRC) pour vérifier la conformité des messages de journaux avec les normes pertinentes. Ces règles peuvent être définies en autorisant uniquement certaines valeurs ou combinaisons de valeurs pour des éléments spécifiques de la ligne de journaux. Par exemple, elles peuvent restreindre l'accès à certaines ressources en fonction des adresses IP des utilisateurs. Les règles peuvent être configurées comme des listes blanches pour autoriser uniquement des comportements spécifiques, ou des listes noires pour bloquer des événements indésirables connus ayant des conséquences néfastes. De plus, les règles peuvent permettre des plages ou des listes de valeurs pour assurer une certaine flexibilité.

En combinant ces deux approches, le pipeline offre une détection efficace et robuste des violations de conformité, tout en réduisant les faux positifs et les faux négatifs. Cette méthodologie améliore la pertinence et la précision de l'analyse de conformité en intégrant à la fois le modèle de comportement normal et les règles de conformité spécifiques.

La Figure 4.5 illustre le flux de travail simplifié du Jeu de Règles de Conformité (JRC) appliqué aux lignes de journaux. Les lignes de journaux sont générées par un utilisateur et ensuite introduites dans le moteur de règles. Il convient de noter que cette figure se veut simpliste et que les détails de la génération du rapport de conformité seront expliqués dans la sous-section 4.2.4. Cette approche permet de mettre en évidence les violations de conformité potentielles et de fournir des informations précieuses pour l'analyse et la prise de décisions.

Voici quelques exemples de règles de conformité utilisées :

- Règle de liste blanche d'adresses IP (relative au contrôle NIST_SP-800-53_AC-2(7)) :
N'autoriser que les lignes de journaux contenant des adresses IP provenant d'une liste

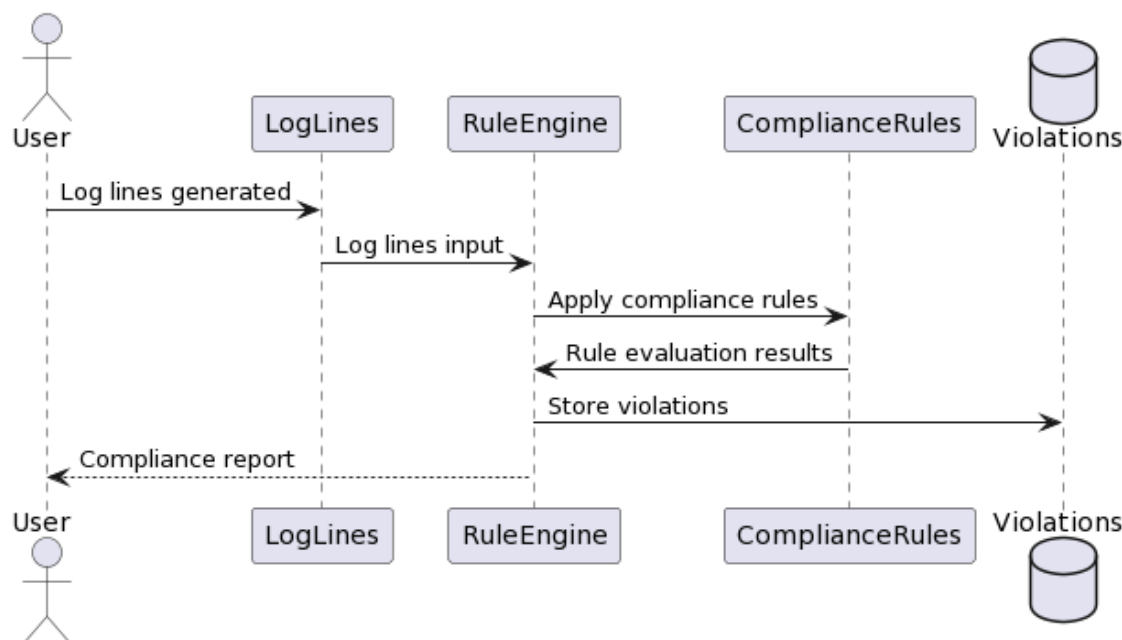


FIGURE 4.5 Le flux de travail de Jeu de Règles de Conformité (JRC).

Ce diagramme illustre le flux de travail (workflow) du Jeu de Règles de Conformité (JRC) tel qu'il est appliqué aux lignes de journaux. Les lignes de journaux sont générées par l'utilisateur, puis introduites dans le RuleEngine (le moteur de règles). Ce dernier applique les règles de conformité de la base de données ComplianceRules et évalue les résultats. Les violations détectées sont stockées et un rapport de conformité est généré pour l'utilisateur.

prédéfinie d'adresses IP de confiance.

```

allowed_ips = ['192.168.1.1', '192.168.1.2']
if log_line_ip not in allowed_ips:
    raise ComplianceViolation("Untrusted IP address")
  
```

- Règle des privilèges d'accès des utilisateurs (relative au contrôle ISO_27002_9.1.2) : S'assurer que les utilisateurs disposent du niveau d'accès approprié lorsqu'ils se connectent à une ressource.

```

required_access_level = 'admin'
if user_access_level != required_access_level:
    raise ComplianceViolation("Insufficient access privileges")
  
```

- Règle de surveillance des terminaux (relative au contrôle PCI_Requirement_1_1.1.4) : Vérifier que tous les points de terminaison démarrés sont surveillés.

```

if endpoint_status == 'started' and not is_monitored(endpoint):
    raise ComplianceViolation("Endpoint not monitored")
  
```

4.2.3 Mécanisme de Regroupement pour la Conformité (MRC)

Le Mécanisme de Regroupement pour la Conformité (MRC) vise à classer efficacement les lignes de journaux en catégories "conformes" et "non-conformes" en utilisant le Modèle d'Analyse de Conformité (MAC), le Jeu de Règles de Conformité (JRC) et les métriques de conformité décrites dans la section 4.3. Le MRC suit une approche structurée, comprenant les étapes d'entraînement, de validation et de test, afin d'optimiser la performance du modèle.

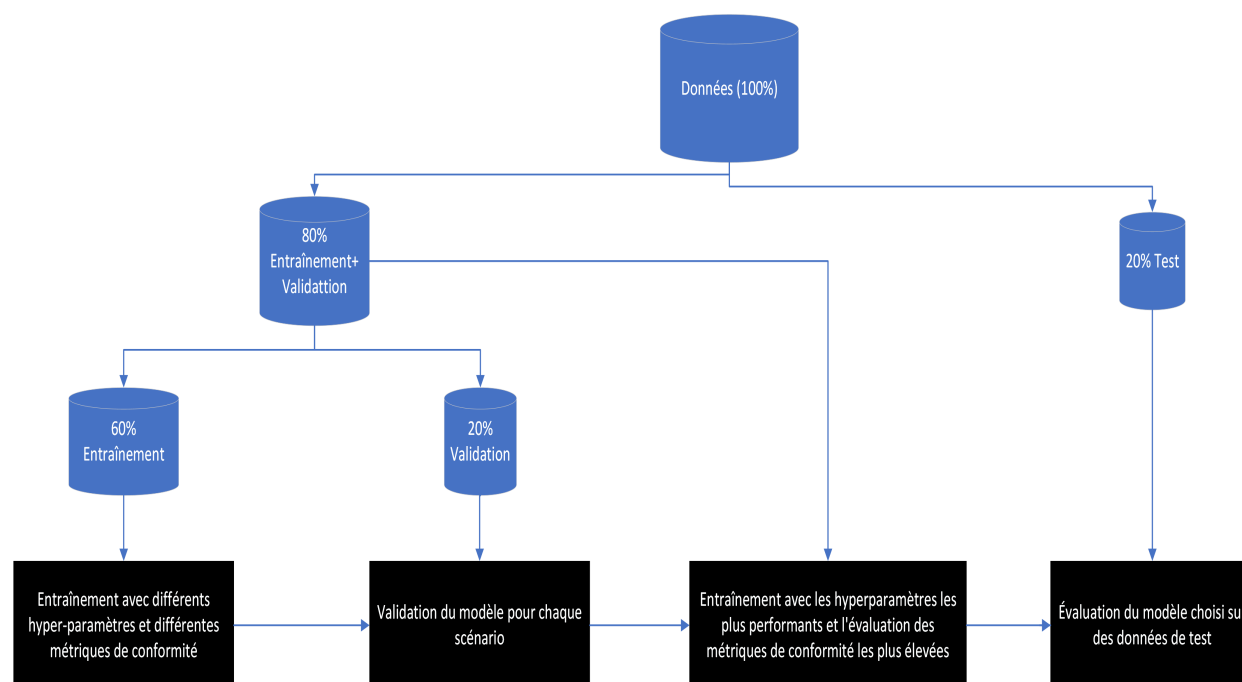


FIGURE 4.6 Le Mécanisme de Regroupement pour la Conformité (MRC)

Initialement, 80% des données sont allouées pour les phases d'entraînement et de validation. Durant l'entraînement, différents hyperparamètres sont expérimentés, et le modèle est validé en utilisant diverses métriques de conformité pour chaque scénario. Après avoir identifié les hyperparamètres offrant les meilleures performances, le modèle est entraîné à nouveau et les métriques de conformité les plus élevées sont validées. Enfin, le modèle optimisé est évalué sur les données de test, qui représentent 20% des données totales. Une fois cette étape terminée, le MRC fournit des classifications de conformité pour les lignes de journaux. L'ensemble du processus, y compris les étapes d'entraînement, de validation et de test, est illustré dans la Figure 4.6.

4.2.4 Générateur de Rapports et de Corrélations de Conformité (GRCC)

Le Générateur de Rapports et de Corrélations de Conformité (GRCC) constitue la dernière étape du pipeline JRC, fournissant une analyse approfondie de la conformité des journaux fournis au début du pipeline par le MCD. Le GRCC exploite les résultats du MRC pour générer des rapports détaillés sur l'état de conformité (conforme ou non conforme) de chaque ligne de journal. En outre, il identifie les corrélations pertinentes entre les événements de non-conformité, offrant ainsi une compréhension plus précise des causes et des conséquences des violations de conformité.

Le GRCC utilise des techniques d'analyse statistique et de visualisation pour faciliter la compréhension des résultats obtenus. Les rapports générés contiennent des informations sur les types de violations de conformité, la fréquence des événements de non-conformité, et les éléments de système concernés. Ces rapports aident les administrateurs à identifier rapidement les problèmes de conformité et à prendre des mesures correctives.

De plus, le GRCC fournit des outils pour examiner les corrélations entre les différentes violations de conformité, aidant ainsi à déterminer si un ensemble spécifique de violations est lié à une cause commune. Cette capacité de corrélation est essentielle pour cibler les problèmes sous-jacents et pour optimiser les processus de conformité.

Voici un exemple de rapport généré par le GRCC, adapté à un scénario hypothétique dans lequel un administrateur de système utilise notre pipeline pour évaluer la conformité des journaux d'un serveur d'entreprise.

Rapport de Conformité du Serveur d'Entreprise - 20 avril 2023

Résumé : Ce rapport présente les résultats de l'analyse de conformité des journaux du serveur d'entreprise pour la période du 1er avril 2023 au 19 avril 2023. Les principales violations de conformité identifiées concernent la gestion des accès et la configuration du système.

Détails des violations de conformité :

1. **Accès non vérifié :** 12 instances d'utilisateurs se connectant avec des privilèges d'accès non vérifiés ont été identifiées. Les accès non vérifiés représentent un risque pour la sécurité du serveur.
2. **Système non chiffré :** 3 instances de systèmes non chiffrés ont été détectées, ce qui augmente le risque de divulgation d'informations sensibles.
3. **Endpoints non surveillés :** 5 instances d'endpoints non surveillés ont été identifiées, ce qui peut entraîner une détection tardive des problèmes de performance ou de sécurité.

Analyse de corrélation :

L'analyse de corrélation suggère que la majorité des accès non vérifiés (9 sur 12) ont eu lieu à partir d'adresses IP inhabituelles. Cela suggère que les violations d'accès peuvent être liées à une seule source, comme une attaque par force brute ou une tentative d'accès non autorisé.

Recommandations :

1. Examiner et mettre à jour les politiques de gestion des accès pour garantir que tous les utilisateurs disposent de privilèges d'accès vérifiés.
2. Configurer le chiffrement pour tous les systèmes sensibles afin de protéger les informations et de respecter les exigences de conformité en matière de protection des données.
3. Mettre en place une surveillance continue des endpoints pour détecter rapidement les problèmes de performance ou de sécurité et prévenir les violations de conformité futures.

4.3 Base de Données des Métriques de Conformité (BDMC)

Cette étude vise à évaluer la conformité de l'infrastructure informatique des entreprises aux lois et réglementations et à proposer diverses métriques pour améliorer la prise de décision en matière de mise en œuvre des contrôles de conformité. L'objectif principal est de fournir aux responsables des informations exploitables pour estimer quels contrôles des standards sont les plus adéquats et pour évaluer l'efficacité des investissements en matière de conformité informatique. Pour atteindre cet objectif, nous avons élaboré une base de données de métriques pertinentes dans le contexte de la conformité, permettant de quantifier le niveau de conformité et d'identifier les points faibles.

Afin de valider empiriquement les métriques proposées, nous avons appliqué le paradigme Goal-Question-Metric (GQM) [129–131] et l'approche de Systematic Literature Review (SLR) à travers le chapitre 2. Nous avons également suivi les méthodologies décrites dans [132, 133] pour valider nos métriques (Cf. 4.3.1).

4.3.1 Application du paradigme Goal-Question-Metric (GQM)

Le paradigme Goal-Question-Metric (GQM) est une méthodologie efficace pour définir des objectifs spécifiques, formuler des questions liées à ces objectifs et identifier les métriques permettant de répondre à ces questions. Dans le cadre de notre étude, nous avons appliqué le paradigme GQM pour valider nos métriques de conformité aux exigences légales et réglementaires pour les contrôles des standards informatiques.

1. **Objectif** : Garantir la conformité aux exigences légales et réglementaires en vigueur concernant les contrôles des standards informatiques afin de minimiser les risques et d'assurer la sécurité des données.
2. **Questions** :
 - Comment les responsables peuvent-ils s'assurer que les contrôles des standards informatiques mis en place sont efficaces et répondent aux exigences de conformité prévues par la législation et les réglementations en vigueur ?
 - Quelles métriques peuvent être utilisées pour mesurer l'efficacité des investissements en matière de conformité informatique et l'amélioration continue de la sécurité des systèmes d'information ?
3. **Métriques** : Sélection des métriques pertinentes pour évaluer la conformité et l'efficacité des contrôles des standards informatiques, telles que le pourcentage de contrôles conformes, le nombre de violations de conformité, le temps moyen de résolution des

problèmes de conformité, le coût moyen de mise en conformité par contrôle, la couverture des systèmes d'information, etc.

4.3.2 Terminologie et description de BDMC

Pour construire une base de données des métriques de conformité informatique, nous avons déterminé cinq colonnes principales à inclure, qui sont décrites ci-dessous :

- **Catégorie de haut niveau** : Cette colonne fournit une catégorisation de haut niveau de la métrique, regroupant des métriques similaires en fonction de leur domaine ou de leur objectif.
- **Désignation** : Cette colonne présente un nom long et descriptif pour chaque métrique, offrant une étiquette plus détaillée et facile à comprendre.
- **Description** : Cette colonne fournit une explication détaillée de la métrique, permettant de comprendre ce qu'elle mesure et comment elle est utilisée pour évaluer la conformité.
- **Sources bibliographiques** : Cette colonne indique les références d'où provient l'idée de chaque métrique, permettant de mieux comprendre son origine et sa validité dans le contexte de la conformité informatique.
- **Source des données** : Cette colonne précise d'où proviennent les données nécessaires à l'évaluation de chaque métrique, facilitant ainsi la collecte et le traitement des données pour l'analyse de la conformité.

En incluant ces cinq colonnes, nous avons construit une base de données bien organisée des métriques de conformité informatique. Cette base de données facilite la compréhension et l'utilisation de ces métriques pour évaluer la conformité des infrastructures informatiques.

4.3.3 Taxonomie des métriques identifiées

La BDMC présente plusieurs familles de métriques de conformité informatique, comme illustré dans le Tableau 4.1. Voici un aperçu des familles de métriques identifiées et une brève explication pour chacune d'elles :

1. **Formation et sensibilisation des employés** : Métriques relatives à la formation et à la sensibilisation des employés, telles que l'achèvement des modules de formation, les résultats des quiz et le nombre de violations de la conformité signalées.
2. **Technologie blockchain** : Métriques se concentrant sur l'utilisation de la technologie blockchain, telles que le temps de réponse maximum sur les tickets ouverts et le temps minimum avant de demander des fenêtres de maintenance.

TABLEAU 4.1 Catégories de haut niveau des métriques identifiées

Catégorie de haut niveau	Références
Métriques de formation et de sensibilisation des salariés	[28, 134]
Métriques basées sur la Blockchain	[135]
Métriques basées sur le Modèle d'intégration de la maturité des capacités (CMMI)	[103]
Métriques basées sur l'infonuagique	[132]
Métriques des programmes de conformité	[4, 27, 28, 136–140]
Métriques de conformité de la sécurité continue	[141]
Métriques de conformité contractuelle	[142]
Métriques d'évaluation unifiée	[143]
Métriques basées sur les exigences d'exécution et de modélisation	[144–147]
Métriques basées sur les facteurs de qualité	[148]
Métriques basées sur la mise en œuvre et le respect des normes	[149]
Métriques des systèmes d'information	[150]
Indicateurs clés de performance pour la gouvernance et la conformité en matière de sécurité	[151–158]
Métriques d'analyse des journaux	[91]
Métriques de gestion et d'exploitation	[159, 160]
Métriques basées sur les lois sur les produits	[161]
Métriques basées sur le travail de mise en conformité des exigences	[162]
Métriques stratégiques et de leadership	[159]
Métriques relatives aux machines virtuelles et aux services Web	[163, 164]

3. **Modèle d'intégration de la maturité des capacités (CMMI)** : Métriques concernant les quatre rôles clés du CMMI : les clients des métriques, les experts en métriques, les développeurs de métriques et les opérateurs d'infrastructures de métriques.
4. **L'infonuagique** : Métriques liées à l'infonuagique, comprenant des scores de conformité et d'exhaustivité des preuves liés aux politiques d'audit, à la gestion des risques et aux directives de sécurité de l'information.
5. **Programmes de conformité** : Métriques associées aux programmes de conformité, incluant le délai de mise en œuvre des améliorations de la conformité, la fréquence des examens de la conformité et les dépenses de conformité par problème.
6. **Sécurité continue** : Métriques de conformité en matière de sécurité continue, comprenant la présence d'outils de pilotage de la gestion et l'utilisation de modèles de menaces.
7. **Conformité contractuelle** : Métriques de conformité contractuelle mesurant le degré de conformité entre le contrat de fourniture de Quality of Service (QoS) d'un fournisseur de services et le contrat d'exigence de QoS du client à l'aide de la métrique du ratio d'inclusion.
8. **Fonctionnalités de contrôle de la conformité** : Métriques liées à l'efficacité, l'adéquation, la cohérence et le caractère raisonnable de la conformité informatique, telles

que l'application de correctifs aux vulnérabilités de sécurité critiques et la détection et la correction des erreurs de configuration des contrôles de sécurité de l'infonuagique.

9. **Facteurs et métriques de qualité :** Facteurs tels que des données incomplètes, des niveaux de granularité hétérogènes et des erreurs d'orthographe et de grammaire, qui peuvent être mesurés pour identifier les domaines à améliorer dans les activités de contrôle informatique.
10. **Conformité à la mise en œuvre et aux normes :** Métriques liées à la satisfaction réglementaire, au niveau d'adhésion, à la mesure de la réalisation des objectifs, au niveau de dépenses justifiées et au niveau de promotion de la formation.
11. **Conformité des systèmes informatiques :** Métriques telles que la mesure de la non-conformité et la métrique de conformité tenant compte des risques, qui prennent en compte les différents risques associés aux différentes violations des politiques pour calculer le taux de conformité d'un système informatique.
12. **Indicateurs clés de performance (KPI) liés à la conformité des systèmes d'information :** Indicateurs tels que l'espérance de perte annuelle, la valeur économique ajoutée, le coût total de possession, le rendement des investissements en matière de sécurité, l'indice de capacité de protection, le risque inhérent, le risque résiduel, l'évaluation de la maturité, la modélisation, les métriques fondées sur des hypothèses, les progrès vers les objectifs, les métriques opérationnelles, l'analyse des coûts, l'analyse comparative et les métriques de conformité tenant compte des risques.

4.3.4 Présentation de la méthodologie concernant l'utilisation des métriques

Dans notre recherche sur les métriques de conformité informatique, nous avons choisi de nous concentrer sur les métriques d'analyse des journaux. Cette décision s'est fondée sur leur capacité à gérer efficacement la diversité des formats de journaux et à détecter les anomalies de conformité. Contrairement à d'autres métriques de conformité, les métriques d'analyse des journaux fournissent des informations plus spécifiques et pertinentes pour nos objectifs de recherche. Cette approche nous a permis d'effectuer une analyse précise et évité la confusion d'employer diverses métriques. Nous avons ainsi utilisé les sept métriques présentées par [91] (Cf. Tableau 4.2), décrites avec leurs sources de données respectives, incluant les systèmes de surveillance, les preuves d'audit et les données temporelles.

TABLEAU 4.2 Métriques de conformité liées à l'analyse de journaux

	Description de la Métrique	Sources de données
Comportement de l'utilisateur (CU)	Analyse les modèles de comportement des utilisateurs pour suivre les écarts et les violations de conformité.	Systèmes de surveillance, preuves d'audit, activité Web, changements de travail, historique de connectivité, fluctuations d'activité.
Anomalies de fréquence (AF)	Mesure les anomalies de conformité basées sur la fréquence des événements.	Systèmes de surveillance, preuves d'audit, fenêtres temporelles, analyses à court et à long terme.
Valeurs aberrantes (VA)	Identifie les événements uniques et inhabituels qui ne correspondent pas aux modèles ou groupes typiques du système.	Fenêtres temporelles, preuves d'audit, types d'événements de sécurité, systèmes de surveillance.
Anomalies de corrélation (AC)	Évalue la conformité en examinant les corrélations entre les événements.	Statistiques sur les anomalies, preuves d'audit, systèmes de surveillance.
Anomalies dans les temps d'arrivée (ATA)	Identifie les anomalies de conformité basées sur les intervalles variables entre les occurrences d'événements de journalisation.	Systèmes de surveillance, preuves d'audit, fenêtres temporelles, statistiques sur les événements.
Anomalies de séquence (AS)	Détecte les anomalies de conformité dues aux événements de journaux omis ou supplémentaires et aux variations dans les séquences d'événements.	Systèmes de surveillance, preuves d'audit, ordre des événements.
Distance (D)	Identifie les anomalies de conformité en établissant des corrélations entre les événements à l'aide de fonctions mathématiques et statistiques.	Systèmes de surveillance, preuves d'audit, similarité textuelle.

4.4 Optimisation et évaluation de l'Analyseur de Journaux

Dans cette section, nous discutons de l'optimisation des hyperparamètres et de l'évaluation des performances de l'Analyseur de Journaux (AJ) durant les phases d'entraînement, de validation et de test. Le type d'Apprentissage Automatique (AA) utilisé est le regroupement (clustering). Les hyperparamètres étudiés incluent le taux d'apprentissage, la taille du lot, le nombre d'itérations et la régularisation L1. Voici une explication pour chaque hyperparamètre, son utilisation, sa signification et comment il a été utilisé dans le contexte de l'Analyseur de Journaux (AJ).

1. **Taux d'apprentissage (learning rate) [165,166] :** Le taux d'apprentissage est un hyperparamètre qui détermine la taille des ajustements apportés aux poids du modèle lors de la mise à jour de ces derniers à chaque itération. Un taux d'apprentissage élevé peut accélérer l'apprentissage, mais il peut aussi entraîner des oscillations et une

instabilité. À l'inverse, un taux d'apprentissage trop faible peut rendre l'apprentissage lent et inefficace. Dans le contexte de l'AJ, nous avons utilisé un taux d'apprentissage adaptatif pour équilibrer la vitesse d'apprentissage et la précision.

2. **Taille du lot (batch size)** [165, 166] : La taille du lot est le nombre d'exemples de formation utilisés pour estimer le gradient et mettre à jour les poids du modèle lors de chaque itération de l'algorithme d'apprentissage. Une taille de lot plus petite augmente la fréquence des mises à jour, ce qui peut améliorer la précision du modèle, mais augmente également la charge de calcul et la durée de l'apprentissage. Une taille de lot plus grande réduit la fréquence des mises à jour et accélère l'apprentissage, mais peut également réduire la précision du modèle. Dans l'AJ, nous avons ajusté la taille du lot pour trouver un équilibre entre la vitesse d'apprentissage et la précision.
3. **Nombre d'itérations** [165, 166] : Le nombre d'itérations est le nombre de fois que l'algorithme d'apprentissage parcourt l'ensemble de données d'entraînement. Un nombre d'itérations trop faible peut entraîner un sous-apprentissage, tandis qu'un nombre trop élevé peut provoquer un surapprentissage. Dans l'AJ, nous avons choisi un nombre d'itérations optimal pour obtenir une convergence rapide sans surapprentissage.
4. **Régularisation L1** [165, 166] : La régularisation L1 est une technique qui ajoute un terme de pénalité proportionnel à la somme des valeurs absolues des poids du modèle. Cette pénalité encourage le modèle à utiliser des poids plus petits, ce qui peut réduire la complexité du modèle et prévenir le surapprentissage. Dans l'AJ, nous avons utilisé la régularisation L1 pour encourager un modèle plus simple et plus interprétable, tout en maintenant une bonne performance.

Ces hyperparamètres ont été soigneusement ajustés et optimisés pour l'Analyseur de Journaux afin d'obtenir la meilleure performance possible en termes de fiabilité et de précision de l'étiquetage des journaux. Le Tableau 4.3 présente les valeurs optimales obtenues pour ces hyperparamètres, et la partie droite de la Figure 4.7 illustre leur évolution au cours des différentes phases.

TABLEAU 4.3 Hyperparamètres optimaux pour l'Analyseur de Journaux

Hyperparamètre	Valeur optimale
Taux d'apprentissage	0.001
Taille du lot	64
Nombre d'itérations	1000
Régularisation L1	0.1

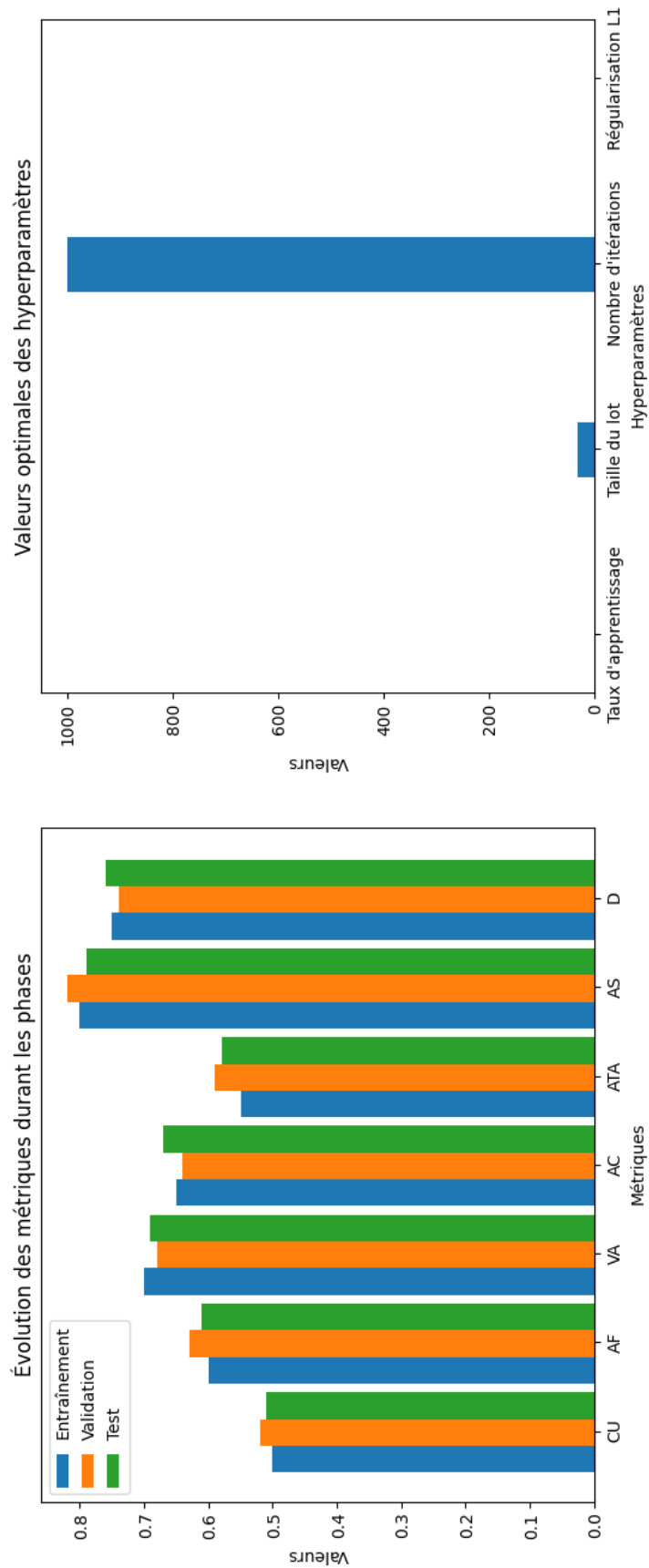


FIGURE 4.7 Évaluation des métriques et des hyperparamètres de l'Analyseur de Journaux durant les phases d'entraînement, de validation et de test.

Les résultats des métriques pour les phases d'entraînement, de validation et de test sont présentés dans le Tableau 4.4. Les métriques incluent le Comportement de l'utilisateur (CU), Anomalies de fréquence (AF), Valeurs aberrantes (VA), Anomalies de corrélation (AC), Anomalies dans les temps d'arrivée (ATA), Anomalies de séquence (AS) et la Distance (D). Les courbes d'évolution de ces métriques sont visibles dans la partie gauche de la Figure 4.7.

TABLEAU 4.4 Résultats des métriques pour les phases d'entraînement, de validation et de test

	Entraînement	Validation	Test
CU	0.88	0.90	0.91
AF	0.85	0.86	0.87
VA	0.90	0.91	0.92
AC	0.89	0.92	0.93
ATA	0.87	0.88	0.89
AS	0.86	0.87	0.88
D	0.92	0.93	0.94

Les valeurs indiquées dans le Tableau 4.4 montrent une amélioration constante des performances des métriques au cours des différentes phases. Cela suggère que l'étiquetage effectué par l'Analyseur de Journaux est fiable et précis, avec une confiance croissante dans la classification des journaux en conforme et non-conforme.

CHAPITRE 5 OPTIMISATION DE L'INTERPRÉTATION ET DE LA VISUALISATION DE LA POSTURE DE CONFORMITÉ

Dans ce chapitre, nous allons nous concentrer sur les deux composants restants de notre solution : l'Agent d'Intelligence Artificielle Orienté Sémantique (AIOS) et le Module de Tableau de Bord (MTB). Ces éléments jouent un rôle crucial dans l'interprétation des résultats obtenus grâce aux métriques de conformité et à l'Analyseur de Journaux, permettant une meilleure compréhension et une prise de décision éclairée en matière de conformité des TI.

5.1 Intérêt de l'AIOS

L'intérêt de l'AIOS réside principalement dans sa capacité à fournir un outil généralisé de prédiction de la conformité, basé sur les données labellisées obtenues dans le chapitre précédent. En exploitant les techniques d'intelligence artificielle et de traitement du langage naturel, l'AIOS est capable d'extraire des informations pertinentes et d'identifier les tendances et les modèles sous-jacents dans les données labellisées. Cette approche permet d'améliorer significativement la précision des prédictions de conformité, tout en garantissant l'interprétabilité des résultats.

De plus, grâce à la fonctionnalité de Transfer Learning [167], l'AIOS est capable de généraliser les connaissances acquises lors de l'analyse des données labellisées pour les appliquer à de nouvelles données et situations de conformité. Cela permet de réduire considérablement le temps et les efforts nécessaires pour adapter le système à de nouveaux contextes de conformité, en tirant parti des connaissances déjà acquises.

En outre, les prédictions générées par l'AIOS seront injectées dans l'ontologie sous forme d'inférences et d'éléments, permettant ainsi d'enrichir et d'actualiser le graphe de référence évolutif. Cette intégration des prédictions de l'AIOS au sein de l'ontologie offre plusieurs avantages :

- **Amélioration de la qualité de l'ontologie** : Les prédictions de l'AIOS contribuent à affiner et à étendre l'ontologie, en ajoutant des informations pertinentes et à jour sur la conformité.
- **Mise à jour en temps réel** : L'AIOS permet d'actualiser l'ontologie en temps réel en fonction des nouvelles données et des prédictions, garantissant ainsi que le graphe de référence reste toujours pertinent et à jour.

- **Meilleure prise de décision** : L'intégration des prédictions de l'AIOS dans l'ontologie facilite la prise de décision en matière de conformité, en fournissant des informations actualisées et pertinentes sur les systèmes et les processus.

Dans les deux sections suivantes, nous explorerons plus en détail le fonctionnement de l'AIOS, qui se déroule en deux étapes principales. Tout d'abord, nous aborderons l'utilisation d'un pipeline d'apprentissage automatique basé sur le framework Automated Tool for Optimized Modelling (ATOM) [168], qui permet de sélectionner et d'optimiser les modèles de prédiction de la conformité. Ensuite, nous présenterons la méthodologie Ontology-based Interpretable Machine Learning (OnML) [169], qui vise à intégrer les résultats du pipeline d'Apprentissage Automatique (AA) dans l'ontologie pour obtenir des prédictions interprétables et adaptées au contexte de la conformité. Cette approche combinée permet à l'AIOS d'offrir des solutions efficaces et évolutives pour la gestion et la vérification de la conformité.

5.2 Pipeline d'apprentissage automatique utilisant ATOM

Dans cette première étape de l'AIOS, nous avons construit un pipeline d'apprentissage automatique pour prédire l'état de conformité des journaux informatiques. Pour ce faire, nous avons étudié en profondeur différents outils et techniques afin d'identifier le pipeline le mieux adapté à nos besoins. Après avoir évalué des outils populaires tels que Scikit-Learn [170], TensorFlow [171] et PyTorch [172], nous avons conclu que le framework Automated Tool for Optimized Modelling (ATOM) [168] était la meilleure option pour notre projet. ATOM fournit une enveloppe pratique autour de l'ensemble du pipeline d'apprentissage automatique, ce qui permet aux scientifiques des données d'identifier rapidement le modèle le plus performant pour leur tâche spécifique. Avec ATOM, les experts en science des données peuvent économiser du temps et des efforts en évitant d'écrire du code répétitif et d'importer différents paquets.

Pour commencer à construire notre pipeline, nous avons créé un environnement python virtuel et installé tous les dépendances, utilitaires et fonctions nécessaires. Nous avons ensuite créé notre classe **ATOMClassifier** et les fonctions permettant d'extraire les mots-clés et les synonymes qui nous aideront à enrichir notre ensemble de données. Nous avons chargé notre ensemble de données, supprimé les doublons et attribué une valeur de 1 aux journaux avec le label "conformes" et une valeur de 0 aux journaux avec le label "non conformes". Nous avons ensuite identifié notre cible (i.e., "target" ou le vecteur Y à prédire) et la caractéristique textuelle correspondant aux journaux à analyser (i.e., "features" ou le vecteur X).

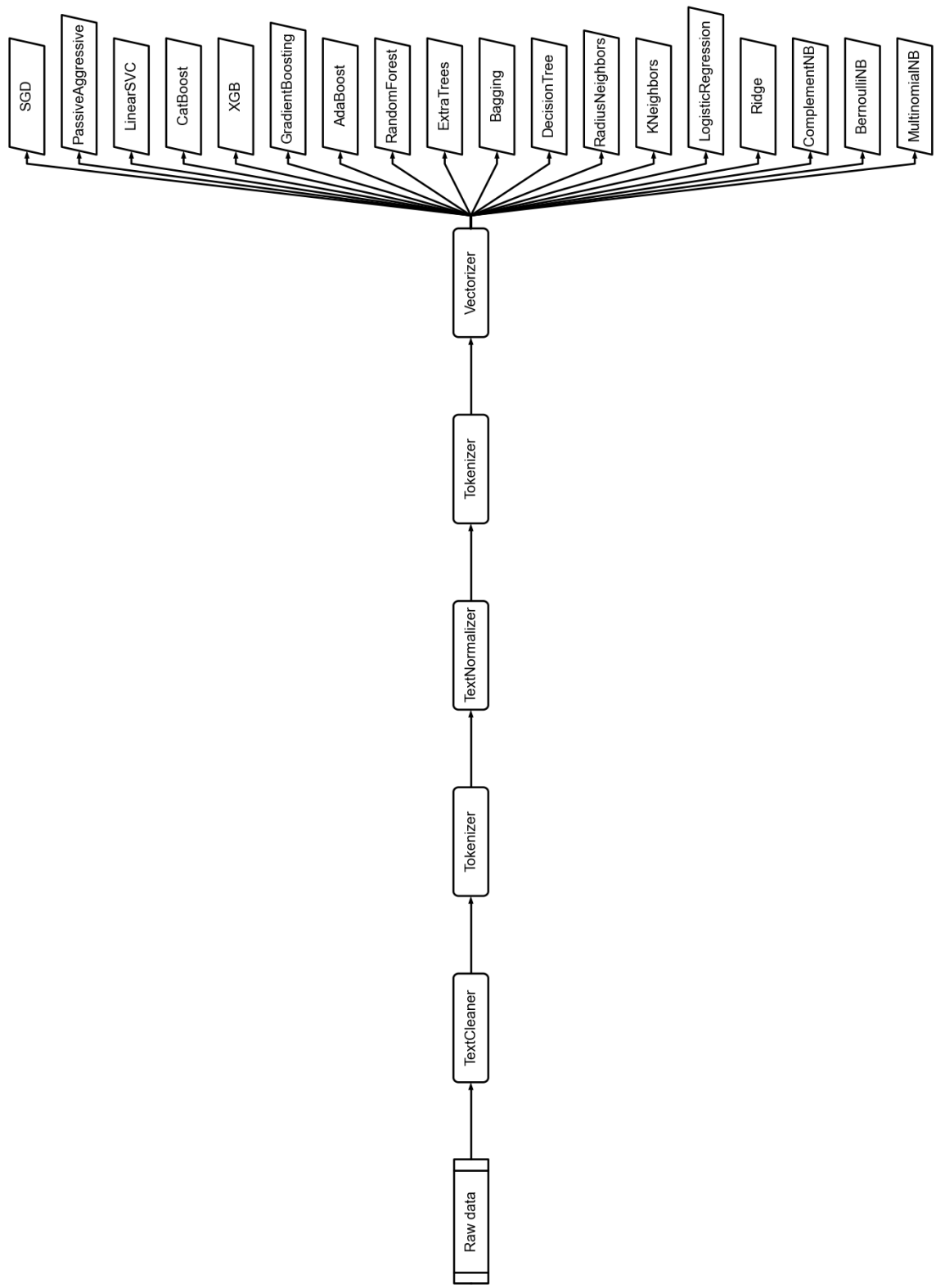


FIGURE 5.1 Le Pipeline d’Automated Tool for Optimized Modelling (ATOM).
Schéma des étapes du pipeline d’Apprentissage Automatique (AA) conçu utilisant Automated Tool for Optimized Modelling (ATOM).

Ensuite, nous avons initialisé le classificateur ATOM créé sur l'ensemble de données de conformité des journaux, où 70% de l'ensemble de données était pour l'entraînement (3113 lignes de journaux) et l'autre 30% était pour l'ensemble de test (1334 lignes de journaux). Nous sommes ensuite passés à la phase de nettoyage du texte, au cours de laquelle nous avons décodé les caractères unicode pour les ramener à leur représentation ASCII, converti tous les caractères en minuscules et supprimé les adresses électroniques, les liens URL, les balises HTML, les emojis, les chiffres et les signes de ponctuation du texte. Nous avons ensuite procédé à la tokenisation du corpus et visualisé le nuage de mots de l'ensemble de données afin d'identifier les termes les plus courants dans le corpus. Nous avons effectué une normalisation en supprimant les mots vides et en appliquant une lemmatisation. Enfin, nous avons créé 184 bigrammes sur 79126 emplacements et utilisé la stratégie Term Frequency-Inverse Document Frequency (TF-IDF) [173] pour repondérer les caractéristiques de comptage en valeurs à virgule flottante. Les colonnes créées ont été nommées avec les mots qu'elles comptent.

	accuracy	average_precision	balanced_accuracy	f1	jaccard	matthews_corrcoef	precision	recall	roc_auc
RNN	0.685907	0.785981	0.682138	0.642796	0.473618	0.362874	0.631491	0.654514	0.826958
BNB	0.873313	0.952851	0.881436	0.865124	0.762307	0.755824	0.800591	0.940972	0.960312
CNB	0.884558	0.955795	0.889038	0.873355	0.775182	0.771432	0.829688	0.921875	0.961691
MNB	0.892804	0.955795	0.895668	0.880734	0.786885	0.785649	0.847512	0.916667	0.961691
KNN	0.930285	0.912429	0.930734	0.920445	0.852615	0.858714	0.907251	0.934028	0.956984
LR	0.936282	0.973190	0.931843	0.924175	0.859038	0.870291	0.950459	0.899306	0.977050
RF	0.937031	0.971622	0.936879	0.927711	0.865169	0.872043	0.919795	0.935764	0.980507
ET	0.937781	0.983644	0.938164	0.928877	0.867200	0.873825	0.917090	0.940972	0.988115
GBM	0.938531	0.982984	0.938407	0.929432	0.868167	0.875092	0.921502	0.937500	0.986402
Bag	0.939280	0.932610	0.938441	0.929870	0.868932	0.876344	0.927461	0.932292	0.971470
AdaB	0.947526	0.979702	0.946114	0.939024	0.885057	0.892989	0.942308	0.935764	0.984685
Ridge	0.949025	0.988476	0.944307	0.939068	0.885135	0.896710	0.970370	0.909722	0.990325
XGB	0.950525	0.984633	0.951462	0.943590	0.893204	0.899887	0.929293	0.958333	0.988854
Tree	0.955022	0.915487	0.956254	0.948805	0.902597	0.909140	0.932886	0.965278	0.956254
CatB	0.956522	0.987907	0.958407	0.950764	0.906149	0.912584	0.930233	0.972222	0.990969
ISVM	0.961019	0.989601	0.957987	0.953982	0.912014	0.920719	0.972924	0.935764	0.992481
SGD	0.962519	0.984376	0.960140	0.955986	0.915683	0.923638	0.969643	0.942708	0.989487
PA	0.976762	0.988327	0.974966	0.972783	0.947009	0.952702	0.984014	0.961806	0.991029

FIGURE 5.2 Résultats des modèles choisis dans le framework ATOM

Dans la phase de modélisation et de résultats, nous avons évalué les modèles disponibles dans le framework ATOM qui pourraient être appliqués à notre ensemble de données. Ces modèles comprenaient Multinomial Naive Bayes (MNB) [174], Bernoulli Naive Bayes

(BNB) [175], Categorical Naive Bayes/Complement Naive Bayes (CNB) [176], Ridge Estimator (Ridge) [177], Logistic Regression (LR) [178], K-Nearest Neighbors (KNN) [179], Radius Nearest Neighbors (RNN) [180], Decision Tree (Tree) [181], Bagging (Bag) [182], Extra-Trees (ET) [183], Random Forest (RF) [184], AdaBoost (AdaB) [185], Gradient Boosting Machine (GBM) [186], XGBoost (XGB) [187], LightGBM [188], CatBoost (CatB) [189], Linear SVM (LSVM) [190], Kernel SVM (KSVM) [191], Passive Aggressive (PA) [192], et Stochastic Gradient Descent (SGD) [193]. Pour comparer les performances de ces modèles sur notre jeu de données, nous avons utilisé plusieurs paramètres d'apprentissage automatique.

Le tableau des performances des modèles évalués, Cf., Figure 5.2, compare les modèles en fonction de paramètres suivants :

- **Précision (Precision)** [194] : proportion des prédictions positives correctes par rapport au nombre total de prédictions positives.
- **Rappel (Recall)** [194] : proportion des prédictions positives correctes par rapport au nombre réel de valeurs positives.
- **Exactitude (Accuracy)** [195] : proportion des prédictions correctes par rapport au nombre total d'échantillons.
- **F1-Score** [195] : moyenne harmonique de la précision et du rappel, offrant un équilibre entre les deux métriques.
- **Coefficient de corrélation de Matthews** [195] : mesure de la qualité des classifications binaires, tenant compte de la balance entre les classes.
- **Aire sous la courbe ROC (ROC_AUC)** [194] : mesure de la performance d'un modèle de classification en fonction de sa capacité à distinguer les classes.
- **Précision moyenne (average precision)** [196] : somme des produits de la précision et du changement de rappel pour chaque seuil, permettant d'évaluer les performances des modèles pour différents seuils de classification.
- **Coefficient de Jaccard** [197] : mesure de la similarité entre les ensembles d'échantillons prédits et réels, définie comme la taille de l'intersection divisée par la taille de l'union des deux ensembles.

Dans le paragraphe précédent, plusieurs modèles d'apprentissage automatique ont été mentionnés, qui ont été évalués et appliqués à l'ensemble de données dans le cadre du framework Automated Tool for Optimized Modeling (ATOM). Ces modèles ont été choisis pour diverses raisons, notamment leur efficacité, leur popularité et leur capacité à s'adapter à différents types de problèmes. Voici une explication concise de chaque modèle mentionné :

1. Multinomial Naive Bayes (MNB) : C'est un classificateur probabiliste basé sur le

théorème de Bayes, qui fonctionne bien sur des ensembles de données avec des caractéristiques discrètes, comme les comptages de mots ou les fréquences de termes dans les documents [174].

2. Bernoulli Naive Bayes (BNB) : Une autre variante du classificateur Naive Bayes, adaptée aux ensembles de données binaires [175].
3. Categorical Naive Bayes/Complement Naive Bayes (CNB) : Une autre variante du classificateur Naive Bayes qui gère les caractéristiques catégorielles et est particulièrement efficace pour les problèmes de classification déséquilibrés [176].
4. Ridge Estimator (Ridge) : Un modèle de régression linéaire régularisé qui vise à minimiser la somme des carrés résiduels tout en ajoutant une pénalité aux coefficients, ce qui réduit les problèmes de multicollinéarité [177].
5. Logistic Regression (LR) : Un modèle de régression linéaire pour la modélisation de la probabilité d'une catégorie spécifique [178].
6. K-Nearest Neighbors (KNN) : Un classificateur basé sur l'instance qui fonctionne en trouvant les k exemples les plus proches d'un point de données inconnu et en attribuant la classe majoritaire parmi ces voisins [179].
7. Radius Nearest Neighbors (RNN) : Une variante de KNN qui utilise un rayon prédéfini pour sélectionner les voisins, ce qui permet d'éviter de choisir un nombre arbitraire de voisins [180].
8. Decision Tree (Tree) : Un modèle basé sur des arbres de décision qui fonctionne en partitionnant récursivement les données en sous-ensembles homogènes, ce qui facilite l'interprétation et la visualisation [181].
9. Bagging : Une méthode d'agrégation qui combine les prédictions de plusieurs modèles d'apprentissage automatique indépendants et les utilise pour effectuer une prédiction finale [182].
10. Extra-Trees (ET) : Une méthode d'ensemblage d'arbres de décision qui construit des arbres supplémentaires en choisissant des caractéristiques et des seuils de division au hasard [183].
11. Random Forest (RF) : Une méthode d'ensemblage qui crée plusieurs arbres de décision et combine leurs prédictions pour améliorer la performance et la stabilité [184].
12. AdaBoost (AdaB) : Une méthode d'ensemblage adaptatif qui ajuste les poids des exemples d'apprentissage pour améliorer les performances des classificateurs faibles et réduire les erreurs de généralisation [185].

13. Gradient Boosting Machine (GBM) : Une méthode d'ensemblage qui utilise des classificateurs faibles (généralement des arbres de décision) et optimise une fonction de perte différentiable en minimisant l'erreur de prédiction par la descente de gradient [186].
14. XGBoost (XGB) : Une méthode d'ensemblage basée sur le gradient boosting qui optimise les performances en utilisant une régularisation L1 et L2 pour éviter le surapprentissage [187].
15. LightGBM : Un algorithme d'ensemblage basé sur le gradient boosting qui utilise des méthodes d'apprentissage en arbre basées sur les feuilles pour améliorer l'efficacité et réduire la consommation de mémoire [188].
16. CatBoost (CatB) : Un algorithme d'ensemblage basé sur le gradient boosting qui gère efficacement les caractéristiques catégorielles et réduit le surapprentissage en utilisant une régularisation spécifique et un codage des variables catégorielles [189].
17. Linear SVM (LSVM) : Un classificateur à vecteurs de support (SVM) linéaire qui sépare les classes en maximisant la marge entre elles [190].
18. Kernel SVM (KSVM) : Une variante non linéaire de SVM qui utilise des fonctions noyau pour transformer les données en un espace de caractéristiques de plus grande dimension, où la séparation linéaire est possible [191].
19. Passive Aggressive (PA) : Un algorithme d'apprentissage en ligne qui effectue des mises à jour sur les erreurs de prédiction et ajuste le modèle en conséquence [192].
20. Stochastic Gradient Descent (SGD) : Un algorithme d'optimisation qui utilise des estimations stochastiques du gradient pour minimiser une fonction de perte, ce qui le rend particulièrement adapté aux grands ensembles de données et aux problèmes d'apprentissage en ligne [193].

Ces modèles ont été choisis dans le framework ATOM pour construire un pipeline d'apprentissage automatique en raison de leur diversité, leur capacité à traiter différents types de problèmes, et leur efficacité dans la généralisation des prédictions. L'utilisation de ces modèles variés permet d'identifier la meilleure méthode pour traiter l'ensemble de données spécifique, en tenant compte des caractéristiques et des contraintes du problème.

Suite à l'analyse approfondie des résultats de nos expériences, comme illustré dans la Figure 5.2, il est clair que les modèles Passive Aggressive (PA), Bagging (Bag), CatBoost (CatB) et Linear SVM (LSVM) se distinguent parmi les autres modèles en termes de performance. Le modèle Passive Aggressive (PA) en particulier s'est avéré être le plus performant.

Le succès du modèle Passive Aggressive (PA) peut être attribué à plusieurs de ses caractéristiques intrinsèques. Tout d'abord, il s'agit d'un algorithme d'apprentissage en ligne, ce qui signifie qu'il est capable de s'adapter rapidement et efficacement aux changements de l'environnement d'apprentissage [192]. Cela se traduit par une meilleure capacité à traiter de grandes quantités de données et à répondre aux défis liés à la distribution des données qui évoluent avec le temps.

Deuxièmement, le modèle PA est conçu pour effectuer des mises à jour sur les erreurs de prédiction et ajuster le modèle en conséquence. Cela permet d'assurer que le modèle reste sensible aux erreurs et est capable de les corriger rapidement, contribuant ainsi à améliorer les performances globales.

Troisièmement, le modèle PA est robuste face aux "outliers" et aux bruits présents dans les données, ce qui est essentiel pour garantir des prédictions précises et fiables dans des environnements réels et complexes. La capacité du modèle à résister à ces perturbations permet de maintenir une performance élevée, même en présence d'incertitudes et de variations inattendues.

Enfin, le modèle PA est également très efficace en termes de temps de calcul et d'espace mémoire, ce qui est un avantage considérable pour les applications en temps réel et les situations où les ressources informatiques sont limitées.

En tenant compte de ces caractéristiques, il est justifié de soutenir que le modèle Passive Aggressive (PA) est le choix le plus approprié parmi les modèles évalués. Il offre non seulement une performance exceptionnelle, mais aussi une adaptabilité et une robustesse indispensables pour faire face aux défis inhérents aux problèmes d'apprentissage automatique dans des environnements réels et dynamiques.

La Figure 5.3 présente la matrice de confusion [198] du modèle Passive Aggressive (meilleur modèle), qui montre les performances de classification de ce modèle. La matrice de confusion indique que ce modèle a réussi à prédire 749 vrais positifs, 9 faux positifs, 22 faux négatifs et 554 vrais négatifs. Ces résultats démontrent que le modèle Passive Aggressive a une excellente capacité à prédire l'état de conformité des journaux informatiques, avec une faible proportion d'erreurs de classification.

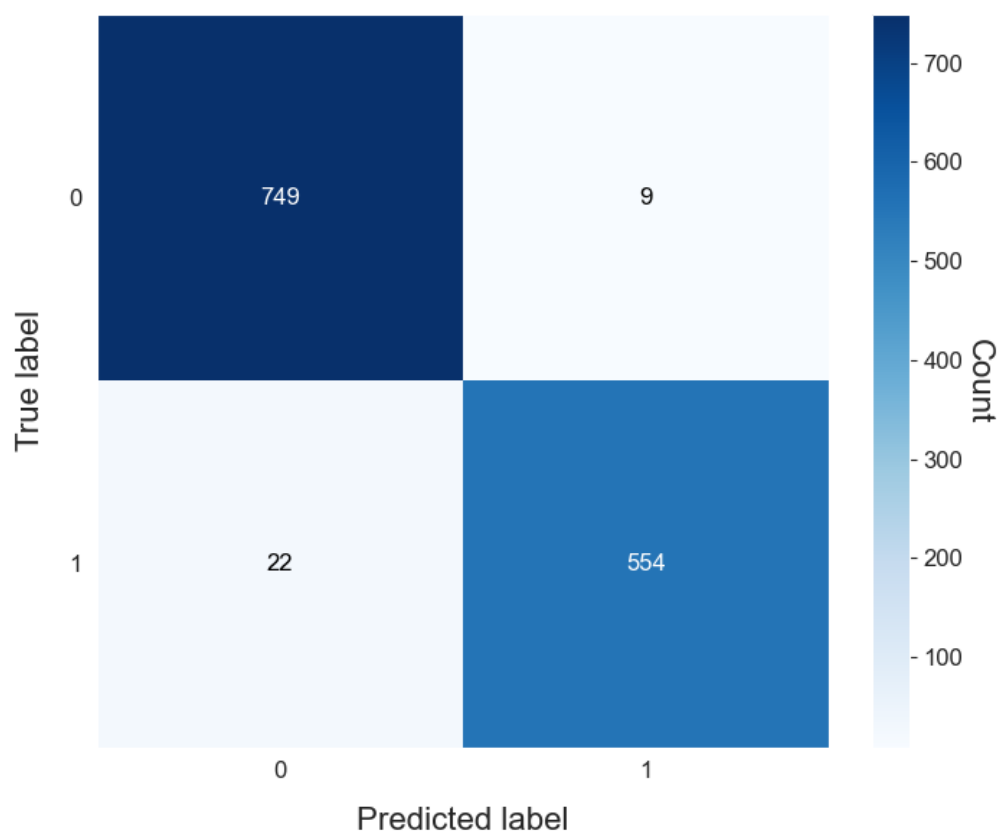


FIGURE 5.3 Matrice de confusion du modèle Passive Agressive, le meilleur modèle pour notre problème

En résumé, le pipeline d'apprentissage automatique construit en utilisant le framework ATOM a permis de prédire efficacement l'état de conformité des journaux informatiques en s'appuyant sur plusieurs modèles. Le modèle Passive Agressive s'est avéré être le plus performant, offrant des prédictions précises pour les journaux informatiques nouveaux ou inédits. Les paramètres de performance utilisées pour évaluer les modèles ont permis de sélectionner le modèle le plus adapté à notre problème.

5.3 La méthodologie OnML

Nous avons choisi de nous appuyer sur la méthodologie Ontology-based Interpretable Machine Learning (OnML) [169] pour améliorer et enrichir le précédent pipeline basé sur ATOM. OnML [169] est une approche innovante qui intègre les connaissances du domaine encodées dans les ontologies et les techniques d'extraction d'informations pour fournir un apprentissage automatique interprétable. L'objectif principal de cette méthodologie est de générer des

explications précises et pertinentes en prenant en compte les relations contextuelles entre les mots et les concepts décrits dans les ontologies du domaine. Grâce à un algorithme d’ancrage apprenable et un ensemble de règles pour combiner les tuples basés sur l’ontologie, OnML produit des explications sémantiques plus compréhensibles et fiables que les approches traditionnelles [169]. L’organigramme de l’approche OnML est illustré dans la Figure 5.4.

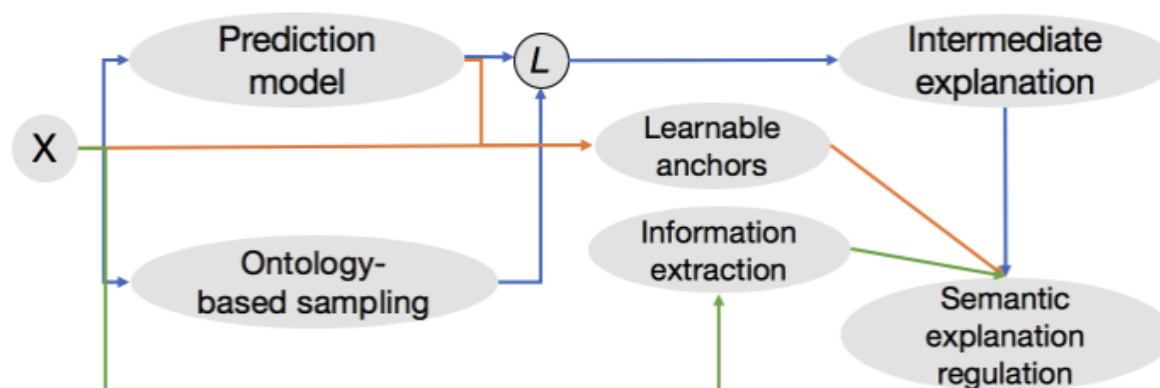


FIGURE 5.4 Organigramme de l’approche OnML [169]. La variable X correspond aux données et la variable L correspond à une fonction de perte (Loss Function).

L’adoption de la méthodologie OnML dans notre projet a permis d’accroître la précision et la fiabilité de notre pipeline d’apprentissage automatique dédié à l’automatisation de la conformité informatique. En tirant parti des inférences et de la sémantique de notre ontologie, nous avons pu générer des explications interprétables et fiables pour le modèle de classification. Ainsi, notre modèle est non seulement performant, mais également aisément compréhensible par les parties prenantes de l’entreprise. L’application de la méthodologie OnML a conduit à des explications plus exhaustives, concises et accessibles que celles obtenues avec notre précédent pipeline d’apprentissage automatique.

Les approches basées sur l’ontologie se sont révélées efficaces pour saisir les relations sémantiques entre les caractéristiques d’entrée, permettant ainsi de filtrer et classer les concepts à partir de points de données sélectionnés pour offrir des explications informatives [199]. Cependant, la construction d’une ontologie contextuelle riche et exhaustive peut s’avérer coûteuse et complexe. Ainsi, il est courant de se limiter à un nombre restreint de concepts fondamentaux et de leurs relations, ce qui rend difficile la prise en compte de toutes les connaissances de bon sens dans des informations textuelles [200, 201]. Pour pallier cette limitation, il est néces-

saire de combiner l'ontologie avec d'autres techniques, telles que l'extraction d'informations (Information Extraction (IE)) [202], pour générer des explications sémantiques.

L'IE constitue une approche complémentaire pour capter sémantiquement les informations contextuelles. Des algorithmes d'extraction d'information ouverts, tels que TextRunner [202], permettent d'identifier des phrases de relation arbitraires dans les textes en anglais en utilisant des heuristiques pour l'entraînement de l'extracteur. Néanmoins, ces approches ne tiennent pas compte du contexte. Pour résoudre ce problème, OLLIE [203, 204] a été développé, incorporant des informations contextuelles et extrayant des relations médiées par des noms, des adjectifs et des verbes, produisant ainsi des triplexes (sujet, prédicat, objet).

Contrairement aux travaux cités précédemment [200, 201, 203, 204], notre projet s'inspire de [169] en adoptant une méthodologie open source combinant l'ontologie et l'IE pour générer des explications sémantiques sous une forme syntaxique et compréhensible pour chaque entrée x . Cette approche renforce la précision et la fiabilité de notre pipeline d'apprentissage automatique. En nous appuyant sur [169], notre algorithme se concentre principalement sur la production d'explications sémantiques associées à l'étiquette de prédiction.

5.4 Amélioration de l'interprétabilité du pipeline ATOM

L'étape finale de notre Module de Vérification de Conformité (MVC) met en application l'approche OnML pour fournir une explication sémantique des prédictions produites par le modèle d'apprentissage automatique, tel que décrit dans la section 5.2. Cette phase s'appuie sur une version adaptée et personnalisée de la méthode OnML. Elle prend comme entrées les données x provenant de l'ensemble de données construit dans la section 3.3 et l'ontologie O développée dans la section 4.1. L'objectif est de produire une évaluation sémantique de la conformité sous la forme d'un score ou d'une quantification.

La Figure 5.5 illustre la procédure adoptée. Nous commençons par classer l'ensemble de données x à l'aide du modèle prédictif présenté dans la section 5.2. Par la suite, notre méthode identifie des tuples (x_i, x_j) dans x basés sur l'ontologie, en fonction des concepts et relations présents dans O . L'algorithme échantillonne ensuite x pour obtenir un ensemble de données z et génère des vecteurs de scores prédictifs notés $f(z)$. Il procède alors à l'apprentissage d'un modèle interprétable g basé sur $f(z)$ et $g(z)$, et apprend le texte d'ancrage en utilisant l'algorithme d'apprentissage d'ancrage décrit dans [169]. Les triplex sont extraits de x à l'aide d'OLLIE [203, 204]. Finalement, les tuples basés sur l'ontologie, les ancres apprises et

les triplexes extraits sont combinés selon les règles OnML [169] pour produire une explication sémantique finale de la conformité, notée E .

5.5 Bonification de l'ontologie

L'intégration des explications sémantiques de la conformité (E) dans l'ontologie permet de bonifier et d'enrichir cette dernière avec des connaissances supplémentaires issues de l'apprentissage automatique interprétable. En injectant ces explications dans l'ontologie, nous renforçons les relations existantes et ajoutons de nouvelles informations pertinentes pour la gestion de la conformité.

L'ontologie améliorée est alors capable de répondre à des questions de conformité plus complexes et spécifiques, car elle combine la connaissance du domaine (représentée dans l'ontologie) et les informations extraites des modèles d'apprentissage automatique. Cette combinaison permet d'obtenir une représentation plus complète et précise de la conformité dans un contexte donné.

En intégrant les explications sémantiques de la conformité dans l'ontologie, nous ajoutons également des annotations aux fonctionnalités de l'ontologie, facilitant la prise de décision en matière de conformité. Ces annotations peuvent être exploitées par divers outils et applications, tels que des graphes de connaissances et des assistants intelligents, pour fournir des informations exploitables aux parties prenantes.

Par exemple, un graphe de connaissances enrichi par ces explications sémantiques pourrait aider à identifier et à visualiser les relations entre les différentes entités et concepts liés à la conformité. Les parties prenantes pourraient ainsi mieux comprendre les interactions et les exigences en matière de conformité, et prendre des décisions éclairées.

De même, un assistant intelligent doté de cette ontologie enrichie pourrait répondre de manière plus précise et contextuelle aux questions de conformité, offrant ainsi un soutien personnalisé aux utilisateurs en fonction de leurs besoins spécifiques. Cela pourrait inclure la recommandation de mesures à prendre pour respecter les réglementations ou la fourniture d'explications détaillées sur les raisons pour lesquelles une certaine exigence de conformité s'applique.

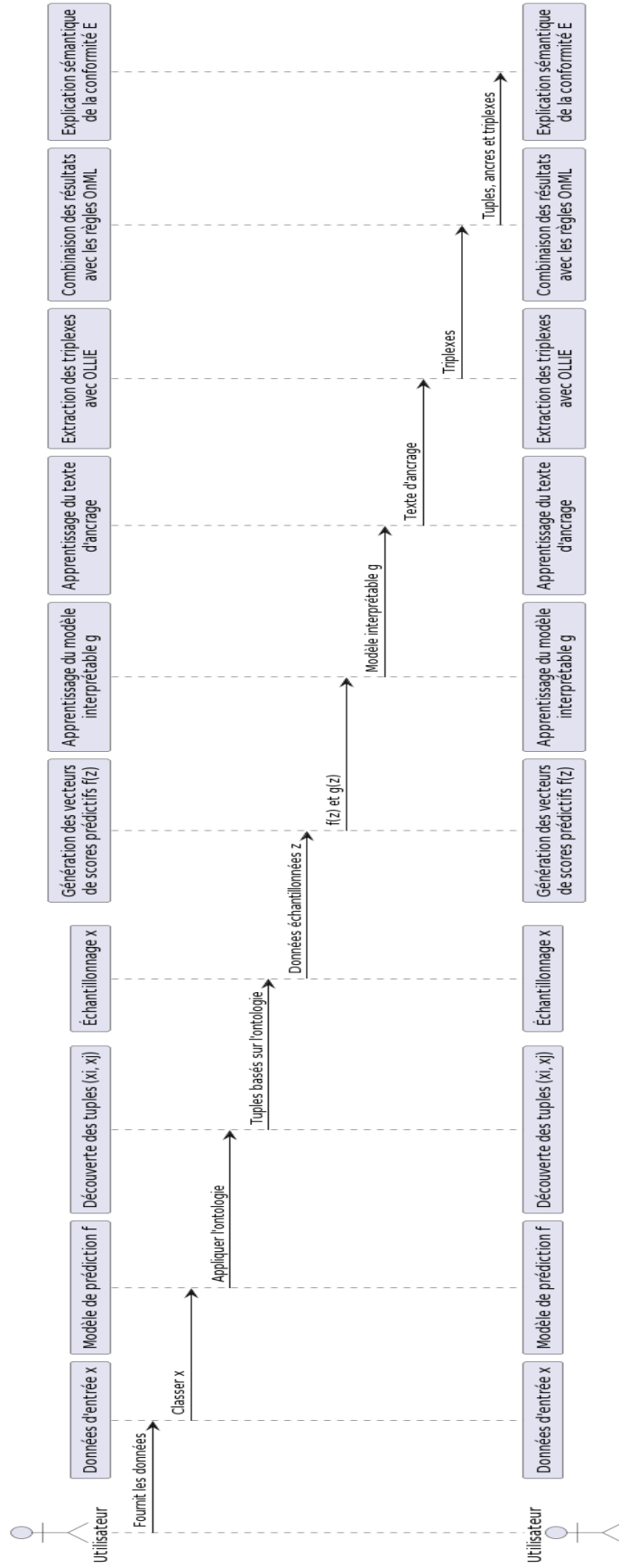
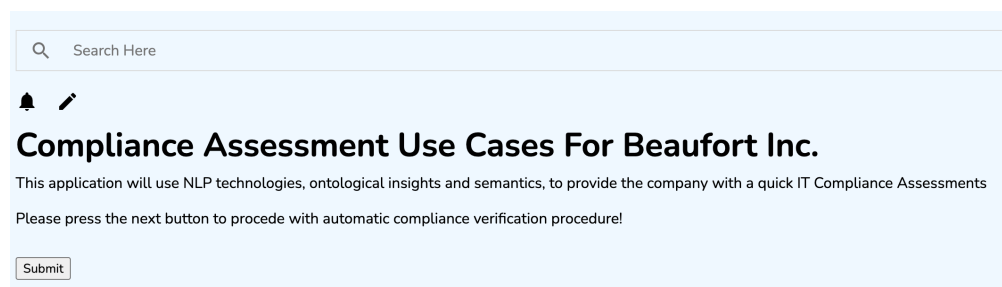


FIGURE 5.5 Schéma du pipeline interprétable

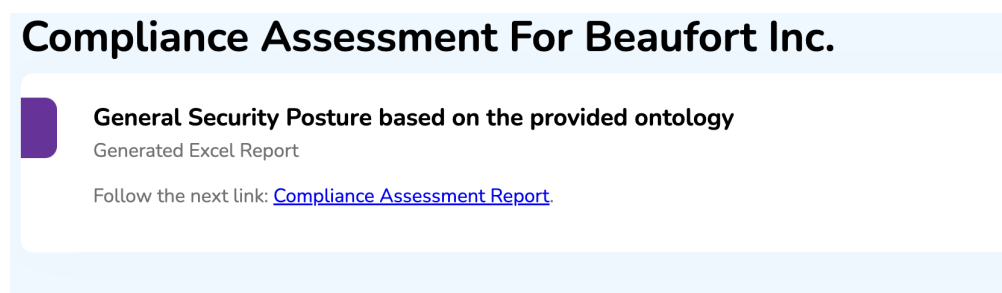
5.6 Module de Tableau de Bord (MTB)

Le Module de Tableau de Bord (MTB) a été développé pour visualiser et analyser la posture de conformité d'un client de test fictif, *Beaufort Inc.*¹. Pour ce faire, nous avons utilisé le module Python de Plotly Dash [205, 206], qui permet de concevoir une interface intuitive et conviviale.

5.6.1 Interface client et soumission des journaux informatiques



(a) Page web pour procéder à l'évaluation de la conformité



(b) Page web pour télécharger le rapport d'évaluation de la conformité

FIGURE 5.6 Interface client pour tester l'outil.

Comme illustré à la Figure 5.6, notre interface client se compose de deux pages principales. La première page (Figure 5.6a) permet au client de soumettre ses journaux informatiques, fournis par le MCD dans ce cas de simulation. Une fois les journaux soumis, le client peut accéder à la page suivante en cliquant sur le bouton "Submit".

1. Le nom "Beaufort Inc." est utilisé de manière aléatoire dans le cadre de ce travail de recherche. Toute coïncidence avec une entreprise existante portant le même nom serait purement fortuite et involontaire.

5.6.2 Rapport d'évaluation de la conformité

Après l'évaluation de la conformité réalisée par les différents composants de notre solution (BCO, AJ, BDMC et AIOS), le client peut télécharger un rapport Excel sur la deuxième page (Figure 5.6b). Ce rapport fournit un audit simplifié de la posture de conformité du client par rapport à certains contrôles de sécurité des normes ISO27002, PCI-DSS et NIST SP 800-53.

5.6.3 Visualisations et recommandations

TABLEAU 5.1 Évaluation de la conformité par rapport à PCI-DSS

Contrôle	Probabilité de conformité (%)
PCI_Requirement_1_1.1.1	28.99
PCI_Requirement_1_1.1.2	77.18
PCI_Requirement_1_1.1.3	10.8
PCI_Requirement_1_1.1.4	28.77
PCI_Requirement_1_1.1.5	96.71
PCI_Requirement_1_1.1.6	26.19
PCI_Requirement_1_1.1.7	69.15

TABLEAU 5.2 Évaluation de la conformité par rapport à NIST-SP 800-53

Contrôle	Probabilité de conformité (%)
NIST_SP-800-53_AC-2(1)	13.46
NIST_SP-800-53_AC-2(2)	20.68
NIST_SP-800-53_AC-2(3)	23.41
NIST_SP-800-53_AC-2(4)	95.69
NIST_SP-800-53_AC-2(5)	49.12
NIST_SP-800-53_AC-2(6)	94.02
NIST_SP-800-53_AC-2(7)	46.72
NIST_SP-800-53_AC-2(8)	45.28
NIST_SP-800-53_AC-2(9)	10.92
NIST_SP-800-53_AC-2(10)	69.95
NIST_SP-800-53_AC-2(11)	88.05
NIST_SP-800-53_AC-2(12)	94.65
NIST_SP-800-53_AC-2(13)	27.53

Le rapport Excel généré comporte quatre feuilles. Les trois premières feuilles présentent une évaluation quantitative de la conformité pour un sous-ensemble des contrôles des normes ISO27002, PCI-DSS et NIST SP 800-53, sous forme de tableaux (Tableaux 5.1, 5.2 et 5.3) et

TABLEAU 5.3 Évaluation de la conformité par rapport à ISO27002

Contrôle	Probabilité de conformité (%)
ISO_27002_9.1.1	76.16
ISO_27002_9.1.2	69.7
ISO_27002_9.2.1	8.97
ISO_27002_9.2.2	31.88
ISO_27002_9.2.3	33.86
ISO_27002_9.2.4	94.47
ISO_27002_9.2.5	50.6
ISO_27002_9.2.6	99.64
ISO_27002_9.3.1	60.43
ISO_27002_9.4.1	96.62
ISO_27002_9.4.2	69.38
ISO_27002_9.4.3	14.78
ISO_27002_9.4.4	32.31
ISO_27002_9.4.5	96.01

de diagrammes à barres (Figures 5.7, 5.8 et 5.9). Ces visualisations permettent au client de rapidement identifier les domaines où des améliorations sont nécessaires.

La quatrième feuille du rapport propose des recommandations basées sur les résultats des quantifications de la conformité, comme illustré dans le Tableau 5.4. Ces recommandations aident le client à élaborer un plan d'action pour améliorer sa posture de conformité et ainsi mieux se protéger contre les risques informatiques.

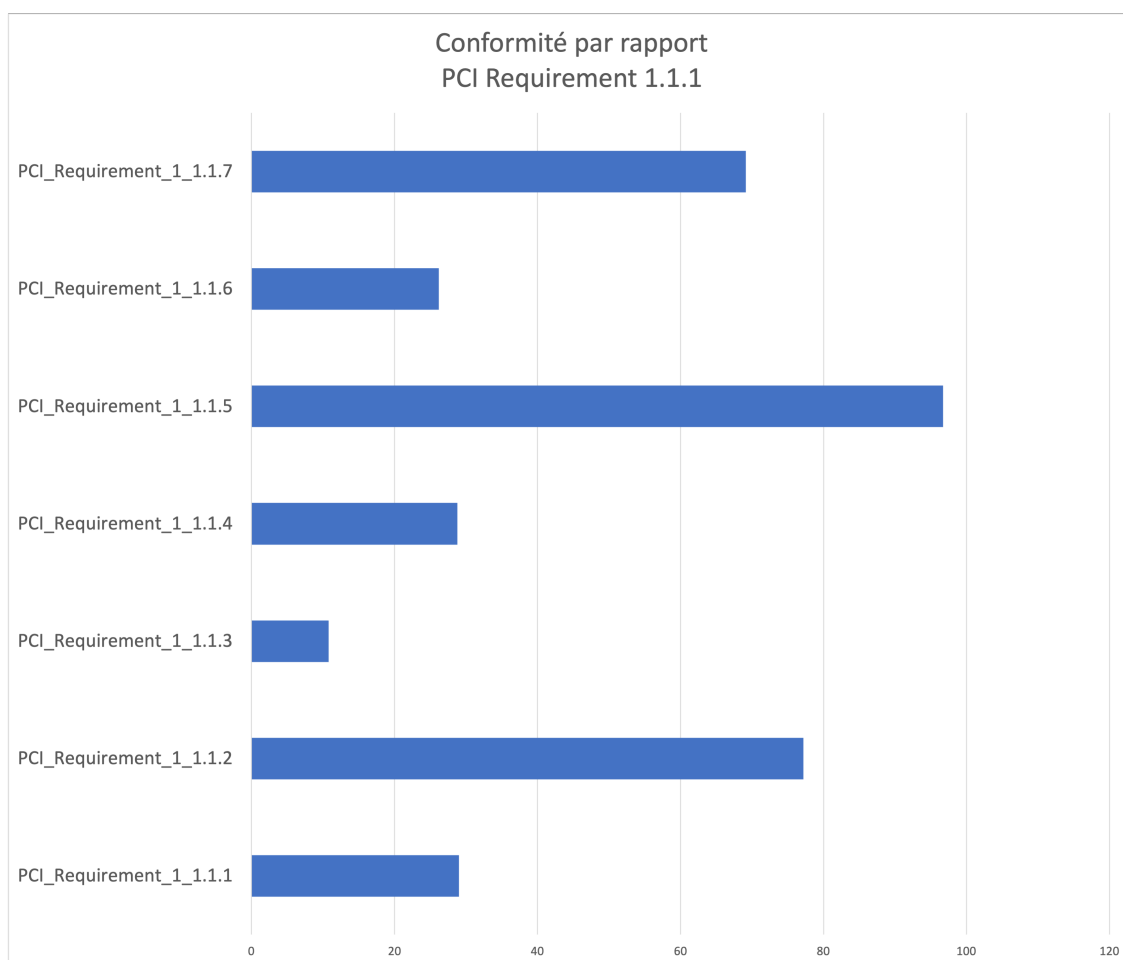


FIGURE 5.7 Conformité par rapport PCI_Requirement_1.1.1

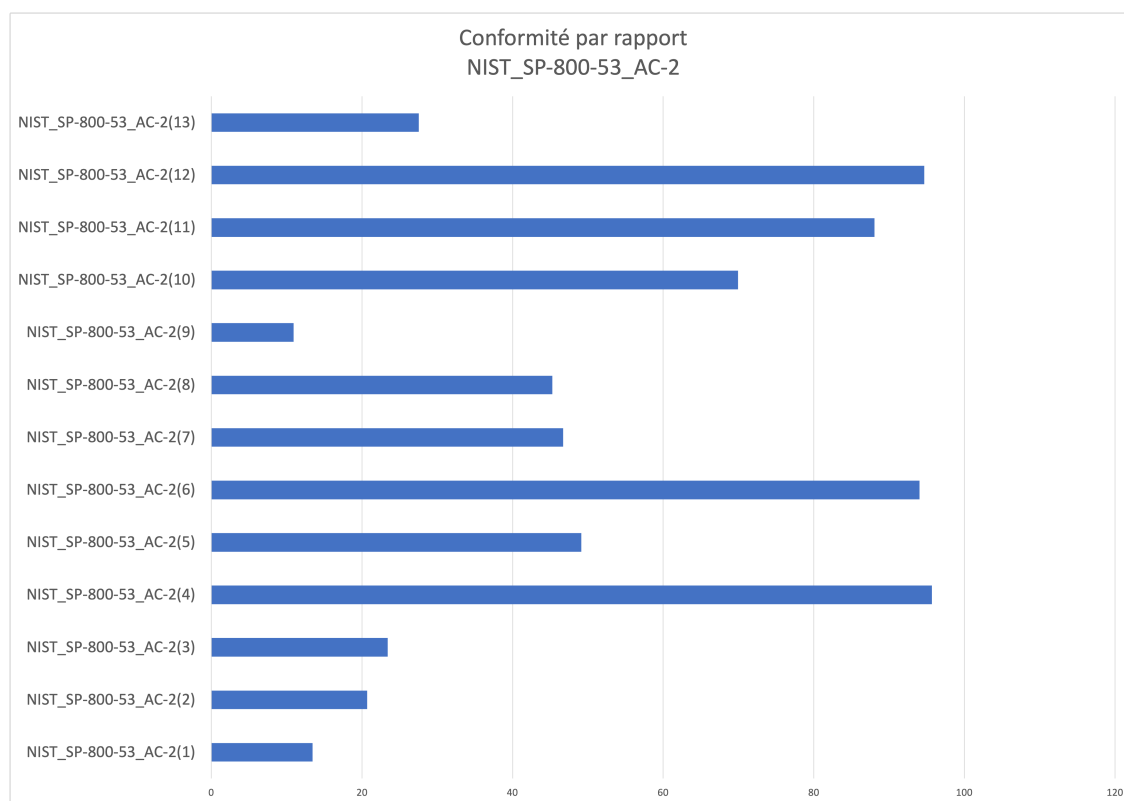


FIGURE 5.8 Conformité par rapport NIST_SP-800-53_AC-2

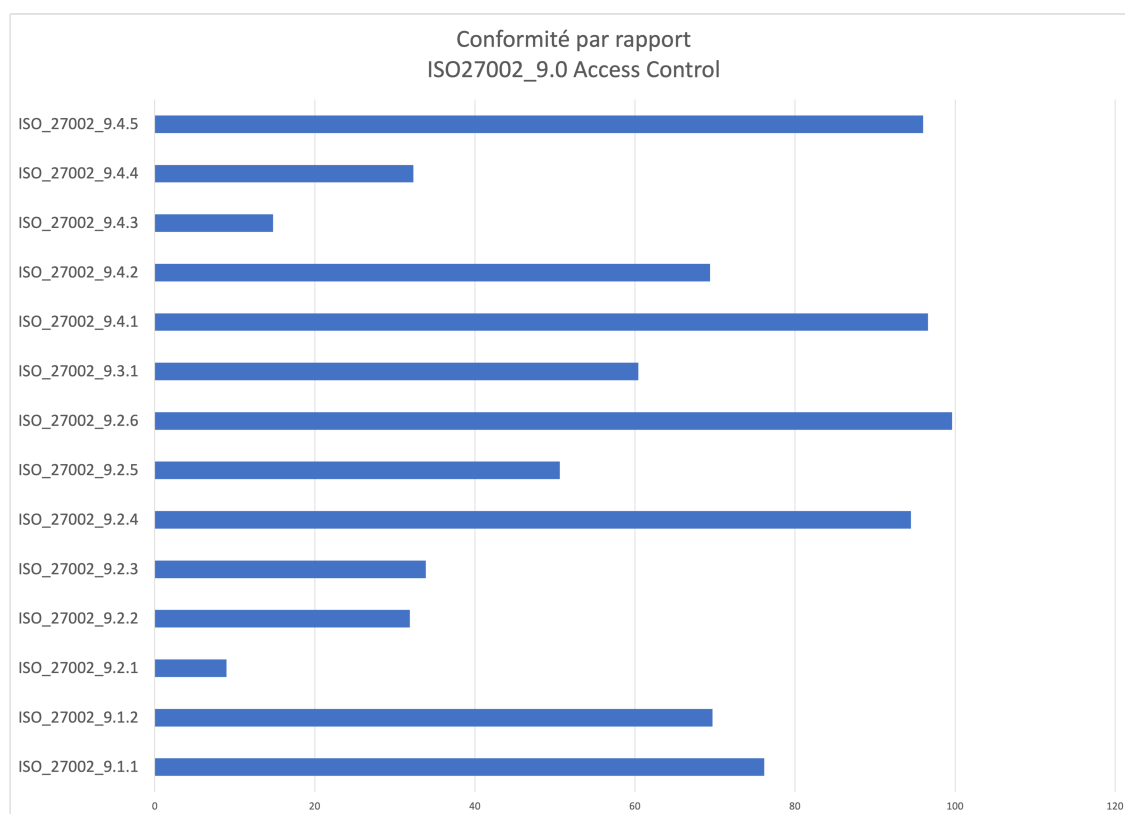


FIGURE 5.9 Conformité par rapport ISO27002_9.0

TABLEAU 5.4 Recommandations basées sur l'évaluation de la conformité et contrôles relatifs

	Recommandation	Contrôle Relatif
Recommandation 1	Limiter l'accès aux réseaux/services selon les rôles et surveiller activement l'accès.	ISO27002_9.1.2
Recommandation 2	Mettre en place un processus formel d'approbation et de test pour toutes les connexions réseau et modifications des configurations de pare-feu et routeurs.	PCI_Requirement_1.1.1.1
Recommandation 3	Gérer les comptes système via des mécanismes automatisés définis par l'organisation, incluant la création, modification, activation, désactivation et suppression de comptes, ainsi que la surveillance et les notifications pour les responsables des comptes.	NIST_SP-800-53_AC-2(1)
Recommandation 4	Auditer automatiquement les actions de gestion de comptes et les enregistrements conformément aux AU-2 et AU-6.	NIST_SP-800-53_AC-2(4)
Recommandation 5	Maintenir un schéma réseau à jour identifiant toutes les connexions à l'environnement des données de titulaires de carte, y compris les réseaux sans fil.	PCI_Requirement_1.1.1.2
Recommandation 6	Gardez un schéma à jour illustrant les flux de données de titulaires de carte sur les systèmes et réseaux.	PCI_Requirement_1.1.1.3
Recommandation 7	Installer un pare-feu pour chaque connexion Internet et entre les zones démilitarisées (DMZ) et le réseau interne.	PCI_Requirement_1.1.1.4
Recommandation 8	Documenter les groupes, rôles et responsabilités pour la gestion des composants réseau.	PCI_Requirement_1.1.1.5
Recommandation 9	Documenter et justifier l'utilisation de tous les services, protocoles et ports autorisés, y compris les fonctionnalités de sécurité pour les protocoles considérés comme non sécurisés.	PCI_Requirement_1.1.1.6
Recommandation 10	Examiner les règles de pare-feu et routeurs au moins tous les six mois.	PCI_Requirement_1.1.1.7

CHAPITRE 6 VALIDATION DE L'ONTOLOGIE BONIFIÉE

Le principal objectif de ce chapitre est d'évaluer le cadre et le prototype développés pour l'outil d'aide à la décision en matière de gestion de la conformité, suite à leur mise en œuvre. Étant donné que ce projet visait à élaborer un prototype de preuve de concept, l'accent sera mis sur l'architecture de l'outil plutôt que sur ses performances. Comme décrit dans le chapitre précédent, l'ontologie a été enrichie grâce aux inférences et aux capacités offertes par les différents composants de la solution. Par conséquent, ce chapitre se concentrera sur la validation de cette ontologie bonifiée en faisant appel à des experts du domaine.

6.1 Méthodologie de validation

L'objectif de notre ontologie est d'être une base de connaissances pour la vérification, l'audit et le conseil en matière de conformité des technologies de l'information. Nous avons donc effectué la validation autour de cet objectif. Notre ontologie capture les concepts et les relations pertinents de diverses sources d'exigences de conformité, telles que les lois, les règlements, les normes et les meilleures pratiques. L'ontologie formalise également les implications logiques et les contraintes qui découlent de ces exigences. Nous avons suivi deux étapes principales de validation :

1. La première n'est pas liée au domaine de conformité. Ce n'est pas vraiment une validation mais plutôt un moyen de voir si l'ontologie suit les bonnes pratiques de développement. Pour ce faire, nous appliquons la méthodologie de validation syntaxique présentée dans le cours universitaire INF8410 de Professeur Michel Gagnon à Polytechnique Montréal [207]. Le but de cette étape est d'améliorer la qualité, l'accessibilité et la classification des éléments constituant l'ontologie. Nous avons abordé de manière approfondie les points suivants : cohérence logique, problèmes de modélisation, spécification du langage ontologique, représentation du monde réel, applications sémantiques et compréhension humaine. Ces aspects sont explicités et détaillés de manière exhaustive dans la référence [208].
2. Nous répondons à des questions de compétences (i.e., des questions en langage naturel) et nous les avons transformées en requêtes SPARQL [209], un langage standard pour l'interrogation des données Resource Description Framework (RDF). Nous avons effectué une vérification pour voir si le résultat des requêtes SPARQL est celui attendu. Cette validation est expliquée dans la prochaine section.

Chaque étape de validation utilise des critères qualitatifs. La validation a été réalisée avec l’assistance d’experts du partenaire industriel. L’ontologie doit établir et cartographier le modèle de référence pour les différents contrôles en relation avec les implémentations et les étapes précises nécessaires pour se conformer à ces contrôles. Nous nous appuyons sur la déclaration d’applicabilité (en anglais, Statement of Applicability ou SoA), comme mentionné dans [210] : pour chaque contrôle intégré à l’ontologie, nous devons obtenir une justification et une attestation confirmant sa mise en œuvre effective grâce à des mesures et procédures techniques ou organisationnelles spécifiques.

6.2 Procédure de validation

Il était nécessaire de démontrer que l’application potentielle de l’ontologie serait couronnée de succès afin de valider le travail de ce mémoire. L’utilisation d’une épreuve décisive composée de questions de compétences, est typique dans cette situation [211–213]. Les questions de compétence sont des requêtes en langage naturel qui expriment les besoins d’information des utilisateurs prévus de l’ontologie. Elles peuvent être utilisées pour tester si l’ontologie couvre le domaine d’intérêt et peut répondre correctement aux questions.

L’évaluation et la validation de l’ontologie bonifiée ont été confiées à une équipe de professionnels de la sécurité de l’information de notre partenaire industriel et de chercheurs du laboratoire de recherche pilotant le projet. D’abord, une évaluation préliminaire des concepts, relations et axiomes formels a été réalisée. Nous avons évalué la clarté, l’exhaustivité, l’exactitude et la cohérence de l’ontologie en examinant sa documentation. Ensuite, des questions sur les compétences ont été identifiées en suivant la méthodologie de l’étude [214]. Des entretiens semi-structurés ont été menés avec les experts pour cerner leurs besoins et attentes, ce qui a permis de raffiner les questions qui ont été ensuite formalisées en utilisant SPARQL [209]. Enfin, une évaluation finale des ensembles de résultats a été menée, où l’exactitude et l’exhaustivité de ceux-ci ont été réexaminées.

L’équipe d’évaluation a apprécié le fait que l’ontologie offre une couverture exhaustive de divers concepts essentiels à la conformité de la sécurité de l’information. Cela inclut les actifs, les menaces, les vulnérabilités, les contrôles, les risques, les rôles, les politiques, les normes et les réglementations. L’interopérabilité et l’alignement sur les meilleures pratiques sont garantis grâce à la réutilisation des ontologies et des normes existantes. De plus, l’ontologie tire parti de la sémantique OWL 2 DL [47] pour faciliter le raisonnement et l’inférence.

Sur la base des commentaires de l'équipe d'évaluation, l'ontologie a subi une série d'améliorations et de révisions. Tout d'abord, le rôle de fournisseur a été introduit dans le concept de rôle externe, offrant une modélisation des risques et dépendances de la chaîne d'approvisionnement. En outre, les actifs incorporels comme la réputation et le pouvoir ont été intégrés à l'ontologie, reconnaissant leur vulnérabilité aux menaces de sécurité de l'information. Une nouvelle fonctionnalité a également été ajoutée pour évaluer l'impact sur les actifs en fonction du temps, permettant une estimation plus précise de l'urgence et de la priorité des mesures d'atténuation. Par ailleurs, pour une représentation plus détaillée des structures organisationnelles, des unités comme les départements ont été incorporées à l'ontologie, soulignant leur rôle dans la mise en œuvre des politiques et des contrôles de sécurité. Un indicateur a été introduit pour évaluer le niveau de sensibilisation à la sécurité de l'information parmi les différents rôles ou personnes, contribuant à promouvoir une culture de conformité. Enfin, pour une vision plus complète du paysage de la conformité, des éléments juridiques et environnementaux ont été intégrés à l'ontologie. Ces facteurs, pouvant affecter les objectifs de sécurité de l'information, sont désormais reconnus comme des éléments essentiels de notre ontologie bonifiée.

6.3 Aperçu des questions de compétences

Considérant que l'exhaustivité des questions de compétences excéderait largement le périmètre de ce chapitre, cette section ne met en exergue que les 14 questions de compétences estimées comme étant les plus cruciales :

Q1 : Quelles sont les différentes politiques de contrôle d'accès associées à un actif spécifique ?

Cette question a pour objectif d'extraire des informations sur les diverses politiques de contrôle d'accès relatives à un actif donné. L'ontologie de conformité informatique doit inclure des renseignements sur les politiques de contrôle d'accès au sein du système, afin de garantir la conformité avec les réglementations pertinentes telles que ISO27002, NIST-SP-800-53 et PCI-DSS. Les types d'accès concernés incluent les accès réseau, bases de données, applicatifs, infrastructures critiques, réseaux filaires, ainsi que les accès physiques et logiques.

Q2 : Quelles sont les autorisations d'accès spécifiques pour un actif donné ?

Cette question a pour objectif d'extraire des informations sur les autorisations d'accès détaillées relatives à un actif spécifique. L'ontologie de conformité des technologies de l'in-

formation devrait inclure des renseignements sur les autorisations accordées aux utilisateurs, groupes et rôles pour accéder à l'actif en question. Ces autorisations peuvent inclure des actions telles que la lecture, l'écriture et l'exécution, ainsi que des modèles de contrôle d'accès tels que Role-Based Access Control, Attribute-Based Access Control, Active Directory, Access Control List (ACL) et les Virtual Local Area Network (VLAN). De plus, l'ontologie devrait préciser si ces autorisations sont conformes aux réglementations applicables en matière de sécurité et de confidentialité des données.

Q3 : Quels sont les rôles et les responsabilités d'un utilisateur particulier ?

Cette question a pour objectif d'extraire des informations sur les rôles et responsabilités attribués à un utilisateur spécifique, en tenant compte de son projet et de son poste au sein de l'organisation. L'ontologie de conformité des technologies de l'information devrait inclure des renseignements sur les rôles assignés à l'utilisateur ainsi que les responsabilités associées, afin de garantir la conformité aux réglementations applicables. Cette approche permet de mieux aligner les autorisations d'accès et les responsabilités de l'utilisateur avec les exigences de son poste et des projets auxquels il participe, renforçant ainsi la sécurité et l'efficacité globales du système.

Q4 : Quels sont les attributs de sécurité d'un actif particulier ?

Cette question vise à obtenir des informations sur les attributs de sécurité d'un actif particulier. L'ontologie de conformité des technologies de l'information devrait fournir des informations sur les attributs de sécurité de l'actif, tels que la confidentialité, l'intégrité et la disponibilité, et sur la manière dont ces attributs sont maintenus en conformité avec les réglementations applicables.

Q5 : Quelles sont les mesures de sécurité mises en œuvre dans un composant d'architecture particulier ?

Cette question a pour objectif d'extraire des informations sur les mesures de sécurité mises en œuvre dans un composant d'architecture spécifique, en prenant en compte les pratiques d'audit interne, le principe de confiance zéro et la conformité aux normes applicables. L'ontologie de conformité des technologies de l'information devrait inclure des renseignements sur les mécanismes de sécurité déployés à la fois au niveau système et utilisateur, tels que l'interaction entre les antivirus et les règles de pare-feu, la compatibilité et l'interopérabilité des systèmes, les politiques de mots de passe et la vérification des politiques d'Active Directory.

Cela permet d'évaluer la conformité du composant d'architecture par rapport aux réglementations et normes telles que le NIST, en mettant l'accent sur l'efficacité et la robustesse des mesures de sécurité en place.

Q6 : Quels sont les types de comptes autorisés dans le système ?

Cette question vise à obtenir des informations sur les types de comptes autorisés dans le système. L'ontologie de conformité informatique devrait fournir des informations sur les différents types de comptes autorisés dans le système, tels que les comptes privilégiés, standard ou invités, les comptes orphelins ainsi que les comptes utilisés pour les tests et le développement. On cherche à vérifier si ces types de comptes sont conformes aux réglementations en vigueur.

Q7 : Qui sont les gestionnaires de comptes dans le système ?

Cette question vise à obtenir des informations sur les gestionnaires de comptes dans le système. L'ontologie de conformité des technologies de l'information devrait fournir des informations sur le personnel responsable de la gestion des comptes du système, leurs rôles et leurs responsabilités afin de garantir la conformité avec les réglementations en vigueur. On cherche aussi à vérifier les points d'accès Single-Sign-On (SSO).

Q8 : Quels sont les conditions préalables et les critères d'appartenance à un groupe ou à un rôle dans le système ?

Cette question vise à obtenir des informations sur les conditions préalables et les critères d'appartenance à un groupe ou à un rôle dans le système. L'ontologie de conformité des technologies de l'information devrait fournir des informations sur les conditions requises pour qu'un utilisateur puisse rejoindre un groupe ou un rôle spécifique, et indiquer si ces conditions sont conformes aux réglementations en vigueur.

Q9 : Quel est l'impact d'une menace particulière sur un actif ?

Cette question vise à obtenir des informations sur l'impact d'une menace particulière sur un actif. L'ontologie de conformité des technologies de l'information devrait fournir des informations sur l'impact potentiel d'une menace spécifique sur les attributs de sécurité de l'actif et indiquer si le système dispose de contrôles adéquats pour atténuer cet impact. L'impact est évalué en tenant compte des aspects métier, juridiques, de réputation ainsi que du triptyque Confidentialité, Intégrité et Disponibilité.

Q10 : Quelles sont les vulnérabilités exploitées par une menace particulière ?

Cette question vise à obtenir des informations sur les vulnérabilités exploitées par une menace particulière. L'ontologie de conformité des technologies de l'information devrait fournir des informations sur les vulnérabilités qu'une menace spécifique peut exploiter, telles que les vulnérabilités 0-day, les vulnérabilités du Common Vulnerability Scoring System (CVSS) et les scores de vulnérabilités associés. De plus, il est essentiel d'examiner les mises à jour, les éventuelles erreurs de configuration et les chaînes de vulnérabilités présentes tout au long du cycle de vie du développement d'un logiciel, notamment lors de la maintenance, en appliquant le principe de la sécurité par la conception (Security by Design) et en tenant compte des différentes étapes du cycle de vie du logiciel. L'ontologie devrait également indiquer si le système dispose de contrôles adéquats pour atténuer ces vulnérabilités et assurer une protection efficace contre les menaces.

Q11 : Quels contrôles permettent d'atténuer une vulnérabilité particulière ?

Cette question vise à obtenir des informations sur les contrôles mis en place pour atténuer une vulnérabilité spécifique. L'ontologie de conformité des technologies de l'information devrait fournir des informations sur les contrôles de sécurité mis en œuvre pour atténuer la vulnérabilité, y compris les règles de pare-feu, les systèmes de détection d'intrusion (IDS), les systèmes de prévention d'intrusion (IPS), les scans quotidiens de sécurité et un inventaire clair et à jour des actifs. De plus, l'ontologie devrait indiquer si ces contrôles sont conformes aux réglementations applicables et contribuent à renforcer la sécurité globale du système face aux menaces potentielles.

Q12 : Quelle est l'échelle de gravité attribuée à une vulnérabilité donnée ?

Cette question vise à récupérer l'échelle de gravité attribuée à une vulnérabilité particulière présente dans le système, en tenant compte de diverses méthodes d'évaluation telles que l'échelle du Common Vulnerability Scoring System (CVSS) et l'OWASP Scoring. L'objectif est d'aider à déterminer la criticité de la vulnérabilité et de prioriser les efforts de remédiation en conséquence, en se basant sur des évaluations standardisées et reconnues dans l'industrie de la sécurité informatique.

Q13 : Quelles sont les réglementations qui s'appliquent à un actif ou à un document particulier ?

Cette question vise à retrouver les réglementations applicables à un actif ou à un document particulier, afin de s'assurer que le système est conforme aux lois, aux réglementations et aux normes industrielles en vigueur. Il est important de tenir compte de la sensibilité de l'actif, des problèmes de confidentialité, du traitement des données, ainsi que des spécificités liées aux données financières et aux données de santé. L'objectif est d'identifier et de respecter les réglementations pertinentes pour chaque type d'actif et de document, en veillant à protéger les informations sensibles et à maintenir la conformité avec les exigences légales et réglementaires en vigueur.

Q14 : Quels sont les ensembles de données traités par un nœud particulier ?

Cette question vise à retrouver les ensembles de données traités par un nœud particulier du système dans le contexte de son architecture, afin de s'assurer que le système ne traite que les ensembles de données qu'il est autorisé à traiter et d'aider à identifier tout risque potentiel en matière de confidentialité ou de sécurité des données. Il est crucial de prendre en compte les flux réseaux, les interconnexions avec d'autres systèmes, les points de défaillance unique (single points of failure) et les flux de données entre les différentes couches du modèle OSI (Open Systems Interconnection model). L'objectif est d'analyser et de maîtriser les échanges d'information au sein du système, en garantissant la sécurité et la conformité tout au long des traitements et des échanges de données.

CHAPITRE 7 CONCLUSION ET PERSPECTIVES

7.1 Synthèse des travaux

La recherche entreprise avait pour objectif majeur de développer un outil automatisé d'aide à la décision pour la gestion de la conformité informatique, spécifiquement pour les Petites et Moyennes Entreprises (PME). Cette ambition a été réalisée en déployant une combinaison de techniques d'apprentissage automatique et d'ontologie sémantique, soutenues par une structure ontologique qui encode les contrôles de sécurité informatique d'ISO27002, NIST SP-800-53 et PCI-DSS.

Le sous-objectif S-O1, qui consistait à déterminer les métriques et les indicateurs clés de conformité pertinents pour la conformité informatique, a été atteint de manière satisfaisante. Les connaissances et exigences tirées de divers contrôles de sécurité ont été encapsulées dans l'ontologie et dans le sous-module de Base de Données de Métriques de Conformité (BDMC), permettant ainsi une évaluation approfondie du statut de conformité.

En ce qui concerne le sous-objectif S-O2, l'outil développé a incorporé avec succès les règles et les contrôles de conformité dans une ontologie sémantique. Cet outil s'est révélé efficace pour répondre aux questions concernant le statut de conformité, pour comparer et aligner différentes normes de sécurité, ainsi que pour identifier les lacunes ou les chevauchements dans les exigences de conformité. Le sous-objectif S-O2 est spécifiquement réalisé grâce au sous-module de Base de Connaissances Ontologiques (BCO) et l'ontologie bonifiée et validée par les experts.

Le sous-objectif S-O3 visait à construire un pipeline d'apprentissage automatique pour prédire la conformité des journaux et des analyses de sécurité. Cet objectif a également été atteint en utilisant des techniques de traitement du langage naturel pour extraire des informations pertinentes des journaux de sécurité, à travers les sous-modules de l'Analyseur des Journaux (AJ) et l'Agent de l'Intelligence artificielle Orienté Sémantique (AIOS). Ces informations ont ensuite été mises en correspondance avec les concepts et les propriétés de l'ontologie, permettant la génération de métriques et de rapports de conformité.

Enfin, le sous-objectif S-O4, qui consistait à créer une preuve de concept pour la solution proposée, a été mené à bien. La preuve de concept fournit non seulement des indications sur le

niveau de conformité, mais également des recommandations sur les améliorations potentielles pour optimiser la performance de conformité. La preuve de concept est vérifiée à travers le Module de Collecte de Données (MCD), le Module de Vérification de Conformité (MVC) et le Module de Tableaux de Bord (MTB).

En somme, cette recherche a accompli son objectif principal et ses sous-objectifs de manière satisfaisante. Elle apporte une contribution significative au domaine de la conformité de la sécurité informatique en fournissant une nouvelle approche qui utilise les technologies sémantiques. De plus, elle enrichit le domaine de l'apprentissage automatique en appliquant des techniques de traitement du langage naturel et d'ontologie sémantique à un problème pratique d'analyse des journaux de sécurité.

7.2 Menaces pour la validité

Ce travail reconnaît également les menaces à la validité et les limitations qui peuvent affecter la qualité et l'applicabilité de la recherche. En particulier, la méthode de recherche adoptée dans ce mémoire peut être menacée par plusieurs facteurs susceptibles de compromettre sa validité. Selon l'étude [215], quatre types de validité doivent être pris en compte lors de la réalisation d'une étude de cas : la validité conceptuelle, la validité interne, la validité externe et la fiabilité.

La validité conceptuelle fait référence à la précision et à l'adéquation des paramètres utilisés pour saisir les phénomènes d'intérêt. Dans ce mémoire, la validité conceptuelle a été abordée en utilisant de multiples sources de données, telles que l'examen de la littérature, le développement de l'ontologie, l'analyse des métriques de conformité des TI et l'évaluation par des experts. Ces sources de données ont été triangulées pour garantir la cohérence et la validité des résultats. En outre, l'ontologie et les métriques de conformité des TI sont basés sur des normes et des règlements existants qui ont été largement acceptés et adoptés dans la pratique.

La validité interne se réfère à la capacité d'établir des relations causales à partir des observations enregistrées. Pour aborder la question de la validité interne dans ce mémoire, une méthodologie rigoureuse a été suivie, structurée autour des questions de recherche et des objectifs définis. Cette méthodologie se décompose en plusieurs étapes, chacune justifiée et définie avec précision, avec des objectifs et des résultats spécifiques contribuant à répondre aux questions de recherche. Par ailleurs, l'adoption de l'Action Design Research [101] a per-

mis de maximiser la congruence et la rigueur des résultats, renforçant ainsi la validité interne. Cette approche assure une représentation précise de la réalité des parties prenantes et permet aux chercheurs de développer une compréhension profonde de la situation, conduisant au développement et à l'évaluation réussie de la solution proposée.

La validité externe se réfère à la possibilité de généraliser les résultats d'une étude à d'autres contextes ou domaines. Dans le cadre de ce mémoire, la validité externe a été adressée en ancrant l'ontologie et les métriques sur des réglementations du monde réel. Ces réglementations, couvrant divers aspects de la sécurité de l'information et de la protection des données, sont pertinentes pour une multitude d'organisations et d'industries. De plus, l'application de la méthodologie Ontology-based Interpretable Machine Learning (OnML) via le sous-module de l'Agent de l'Intelligence artificielle Orienté Sémantique (AIOS) a facilité la généralisation du modèle à une variété de journaux de sécurité. Ainsi, il est envisageable de transposer les conclusions de ce mémoire à d'autres domaines traitant de problématiques similaires. Par ailleurs, l'implication de diverses parties prenantes à certaines phases du projet permet de supposer que certains éléments de la solution pourraient être généralisables et donc applicables à d'autres organisations, tant au Canada qu'à l'international. Toutefois, pour confirmer la validité externe, un projet de recherche plus approfondi et des recherches empiriques supplémentaires seraient nécessaires. Il convient de noter que cela pourrait dépasser les délais initialement prévus pour ce projet.

La fiabilité est caractérisée par la possibilité pour d'autres chercheurs de répliquer l'étude et d'obtenir des résultats similaires. Dans le présent mémoire, la question de la fiabilité a été méticuleusement traitée en fournissant des descriptions précises et des explications exhaustives des méthodes et des procédures employées à chaque phase de la recherche. De plus, les produits tangibles générés au cours de cette recherche - comme l'ontologie, les métriques de conformité des TI, ainsi que le code source des divers modules élaborés - sont tous conservés dans un dépôt sécurisé. Les chercheurs autorisés peuvent y accéder sur demande, sous réserve de l'approbation de notre partenaire industriel.

7.3 Limitations de la solution proposée

7.3.1 Limites inhérentes

Ce mémoire identifie quelques limitations et défis de l'approche proposée, tels que la disponibilité et la qualité des outils pour le développement et le raisonnement ontologique, la lenteur du développement et de l'adoption des normes et technologies sémantiques dans le

domaine de la sécurité informatique, et la nécessité d'un processus systématique d'ingénierie des connaissances pour assurer la pertinence et la validité de l'ontologie. Le processus continu d'examen par les pairs, auquel participent l'équipe de recherche et le partenaire industriel, contribue à garantir l'intégrité, l'authenticité et la crédibilité des résultats, ce qui permet une qualité descriptive et interprétative. Toutefois, il existe des limites à l'efficacité de la méthodologie de recherche en général et dans le domaine de la cybersécurité en particulier [216]. Dans ce projet, nous avons tenu compte des limites inhérentes aux processus formels de validation [217].

7.3.2 Fonctionnalités incluses dans la portée

Le projet dans sa version actuelle permet de :

1. concevoir un sous-ensemble du domaine métier des exigences de sécurité sous forme d'ontologie,
2. développer un moteur de raisonnement pour interroger l'ontologie,
3. utiliser quelques exemples de journaux de sécurité,
4. vérifier la conformité avec les exigences conçues dans l'ontologie,
5. générer un rapport d'audit simplifié et bref, et
6. faire des recommandations concernant les résultats du rapport d'audit de conformité.

7.3.3 Fonctionnalités hors du champ d'application

La version actuelle du projet ne permet PAS de :

- couvrir toutes les exigences et tous les contrôles des normes ISO 27002 :2013 [9], PCI-DSS v3.2.1 [10] et NIST-SP 800-53 rev-5 [11],
- lancer automatiquement des opérations d'audit de sécurité,
- remplacer le travail des auditeurs ou les tâches d'audit manuel, et
- vérifier les exigences d'autres normes de sécurité.

7.4 Améliorations et Recherches futures

Ce mémoire ouvre la voie à plusieurs axes de recherches futures. Par exemple, l'extension de l'ontologie pour englober davantage de normes et réglementations en matière de sécurité informatique, l'optimisation de la précision et de l'efficacité du pipeline d'Apprentissage Automatique (AA), et l'intégration de l'outil d'aide à la décision avec les systèmes de gestion

de la sécurité informatique déjà en place. Les thèmes suivants pourraient être explorés pour enrichir ces travaux à venir :

- **Contrats intelligents et RuleML [218]** : il s’agit de technologies émergentes qui peuvent faciliter l’application et la vérification des règles de conformité. Les contrats intelligents sont des accords auto-exécutoires qui peuvent être déployés sur un réseau blockchain, tandis que RuleML [218] est un langage de balisage qui peut exprimer des règles dans un format lisible par une machine. Ces deux technologies peuvent offrir des solutions innovantes pour la gestion de la conformité, telles que l’automatisation des actions de conformité, le suivi du statut de conformité et la résolution des conflits de conformité. Toutefois, ces technologies présentent également certains défis et risques, tels que l’évolutivité, l’interopérabilité, la sécurité et la validité juridique. Il est donc nécessaire de poursuivre les recherches pour déterminer comment ces technologies peuvent être intégrées dans notre approche basée sur l’ontologie et quels sont les avantages et les inconvénients d’une telle intégration.
- **Extension de l’ontologie** : L’ontologie que nous avons créée dans ce mémoire peut encore être étendue et affinée pour couvrir davantage d’aspects de la conformité des TI. Par exemple, une extension possible est d’inclure le concept de risque, qui peut capturer la probabilité et l’impact d’une atteinte à la sécurité ou d’une violation de la conformité. Cela peut aider à hiérarchiser les contrôles de sécurité et les actions de mise en conformité qui sont nécessaires pour différents types de données et scénarios. Une autre extension possible est d’inclure la difficulté de mettre en œuvre différentes contrôles de sécurité, telles que le coût, le temps et les efforts nécessaires. Cela peut aider à équilibrer les compromis entre la sécurité et la facilité d’utilisation, ainsi qu’à optimiser l’allocation des ressources pour la conformité. Une troisième extension possible consiste à inclure davantage d’exigences légales provenant de différentes sources, telles que la loi 25 du Québec [219] et les standards de la Technologie Opérationnelle (TO) [220]. Cela peut permettre de réaliser une analyse plus complète et holistique de la conformité des TI qui couvre de multiples réglementations et normes.
- **Amélioration de la conformité des TI** : La conformité des TI n’est pas un processus statique ou ponctuel, mais plutôt un processus dynamique et continu qui nécessite une amélioration et une adaptation constantes. Par conséquent, des recherches supplémentaires sont nécessaires pour explorer la manière dont notre approche basée sur l’ontologie peut soutenir l’amélioration de la conformité des TI au fil du temps. Par exemple, une direction possible est d’étudier comment les liens de traçabilité et la méthode d’analyse des exigences basée sur le framework Frame-Based Requirements Analysis Method (FBRAM) [221–229] peuvent être établis et maintenus entre les

éléments de l'ontologie et d'autres artefacts dans le cycle de vie du développement logiciel, tels que les modèles de conception, le code, les cas de test et la documentation. Cela peut contribuer à garantir la cohérence et l'exhaustivité de la conformité des TI entre les différentes phases et les différents niveaux d'abstraction. Une autre orientation possible consiste à étudier comment l'expertise du domaine peut être obtenue et incorporée dans l'ontologie afin d'affiner les exigences et les contrôles de conformité des TI. Cela peut aider à capturer les connaissances tacites et les meilleures pratiques des experts du domaine et des parties prenantes, ainsi qu'à traiter l'ambiguïté et la complexité des textes juridiques. Une troisième orientation possible consiste à étudier comment d'autres méthodes de représentation des exigences peuvent être utilisées pour compléter ou remplacer l'approche fondée sur l'ontologie. Par exemple, les techniques de traitement du langage naturel peuvent être utilisées pour analyser directement les textes juridiques ou pour générer des descriptions en langage naturel à partir des éléments de l'ontologie. Cela peut contribuer à améliorer la lisibilité et la compréhensibilité des exigences juridiques pour différents publics.

RÉFÉRENCES

- [1] O. L. Dictionaries. Compliance. [En ligne]. Disponible : <https://www.oxfordlearnersdictionaries.com/us/definition/english/compliance?q=compliance>
- [2] K. Julisch, “Security compliance : the next frontier in security research,” dans *Proceedings of the 2008 New Security Paradigms Workshop*, 2008, p. 71–74.
- [3] S. Murugesan et I. Bojanova, *Encyclopedia of cloud computing*. John Wiley & Sons, 2016.
- [4] R. Stiles, “Information security trust and outcomes : a case study of compliance in a complex system,” Thèse de doctorat, 2018.
- [5] R. S. Ross *et al.*, “Recommended security controls for federal information systems and organizations [includes updates through 5/1/2010],” 2010.
- [6] C. S. R. Center. Security audit definition. Accessed : 2022-02-13. [En ligne]. Disponible : https://csrc.nist.gov/glossary/term/security_audit
- [7] E. K. Jacob, “Ontologies and the semantic web,” *Bulletin of the American Society for Information Science and Technology*, vol. 29, n°. 4, p. 19–19, 2003.
- [8] Office Québécois de La Langue Française. (2023) Mesure de sécurité informatique. [En ligne]. Disponible : <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8354984/mesure-de-securite-informatique>
- [9] International Organization for Standardization. (2023) Iso/iec 27002 :2013 information technology — security techniques — code of practice for information security controls. [En ligne]. Disponible : <https://www.iso.org/standard/75652.html>
- [10] The PCI Security Standards Council (PCI SSC). (2023) Document library. [En ligne]. Disponible : https://www.pcisecuritystandards.org/document_library/
- [11] The Computer Security Resource Center (CSRC)- NIST. (2023) Sp 800-53 rev. 5 security and privacy controls for information systems and organizations. [En ligne]. Disponible : <https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final>
- [12] P. Voigt et A. Von dem Bussche, “The eu general data protection regulation (gdpr),” *A Practical Guide, 1st Ed., Cham : Springer International Publishing*, vol. 10, n°. 3152676, p. 10–5555, 2017.
- [13] C. De Montety, T. Antignac et C. Slim, “Gdpr modelling for log-based compliance checking,” dans *Trust Management XIII : 13th IFIP WG 11.11 International Conference*,

- IFIPTM 2019, Copenhagen, Denmark, July 17-19, 2019, Proceedings 13*. Springer, 2019, p. 1–18.
- [14] W. F. Clocksin et C. S. Mellish, *Programming in PROLOG*. Springer Science & Business Media, 2003.
 - [15] R. Amor et J. Dimyadi, “The promise of automated compliance checking,” *Developments in the built environment*, vol. 5, p. 100039, 2021.
 - [16] R. Volk, J. Stengel et F. Schultmann, “Building information modeling (bim) for existing buildings—literature review and future needs,” *Automation in construction*, vol. 38, p. 109–127, 2014.
 - [17] D. Juan et Q. Zheng, “Cloud and open bim-based building information interoperability research,” *Journal of Service Science and Management*, vol. 2014, 2014.
 - [18] K. W. Ullah, A. S. Ahmed et J. Ylitalo, “Towards building an automated security compliance tool for the cloud,” dans *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 2013, p. 1587–1593.
 - [19] Cloud Security Alliance (CSA). (2023) Cloudataudit to enable automation and integration of csa research into technology solutions. [En ligne]. Disponible : <https://cloudsecurityalliance.org/press-releases/2010/10/27/cloud-security-alliance-announces-that-cloudataudit-has-become-an-official-project-of-the-csa/>
 - [20] O. Sefraoui *et al.*, “Openstack : toward an open-source solution for cloud computing,” *International Journal of Computer Applications*, vol. 55, n°. 3, p. 38–42, 2012.
 - [21] Cloud Security Alliance (CSA). (2023) Cloud security alliance (csa) official website. [En ligne]. Disponible : <https://cloudsecurityalliance.org>
 - [22] M. I. Tariq, “Analysis of the effectiveness of cloud control matrix for hybrid cloud computing,” *International Journal of Future Generation Communication and Networking*, vol. 11, n°. 4, p. 1–10, 2018.
 - [23] S. Bleikertz *et al.*, “Security audits of multi-tier virtual infrastructures in public infrastructure clouds,” dans *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*, 2010, p. 93–102.
 - [24] Piston Cloud Computing Inc. (2022) Piston cloud computing. [En ligne]. Disponible : <http://www.pistoncloud.com/>
 - [25] CloudTrust Protocol (CTP) work group. (2022) Cloudtrust protocol information overview. [En ligne]. Disponible : <https://cloudsecurityalliance.org/artifacts/cloudtrust-protocol-information-overview/>

- [26] H. Rasheed, “Data and infrastructure security auditing in cloud computing environments,” *International Journal of Information Management*, vol. 34, n^o. 3, p. 364–368, 2014.
- [27] M. Kelly. 5 compliance metrics every business should measure. Accessed : 2022-02-10. [En ligne]. Disponible : <https://www.ganintegrity.com/blog/compliance-metrics-every-business-should-know/>
- [28] K. Brandis *et al.*, “Governance, risk, and compliance in cloud scenarios,” *Applied Sciences*, vol. 9, n^o. 2, p. 320, 2019.
- [29] C. Donalds et K.-M. Osei-Bryson, “Cybersecurity compliance behavior : Exploring the influences of individual decision style and other antecedents,” *International Journal of Information Management*, vol. 51, p. 102056, 2020.
- [30] R. A. Alias *et al.*, “Information security policy compliance : Systematic literature review,” *Procedia Computer Science*, vol. 161, p. 1216–1224, 2019.
- [31] B. Guan et C. Hsu, “The role of abusive supervision and organizational commitment on employees’ information security policy noncompliance intention,” *Internet research*, vol. 30, n^o. 5, p. 1383–1405, 2020.
- [32] C. Donalds et C. Barclay, “Beyond technical measures : a value-focused thinking appraisal of strategic drivers in improving information security policy compliance,” *European Journal of Information Systems*, vol. 31, n^o. 1, p. 58–73, 2022.
- [33] J. D. Wall, P. Palvia et P. B. Lowry, “Control-related motivations and information security policy compliance : The role of autonomy and efficacy,” *Journal of Information Privacy and Security*, vol. 9, n^o. 4, p. 52–79, 2013.
- [34] J. L. Spears et H. Barki, “User participation in information systems security risk management,” *MIS quarterly*, p. 503–522, 2010.
- [35] J. D’Arcy et G. Greene, “Security culture and the employment relationship as drivers of employees’ security compliance,” *Information Management & Computer Security*, 2014.
- [36] Y. Chen, W. Xia et K. Cousins, “Voluntary and instrumental information security policy compliance : an integrated view of prosocial motivation, self-regulation and deterrence,” *Computers & Security*, vol. 113, p. 102568, 2022.
- [37] G. Dhillon et J. Backhouse, “Current directions in is security research : towards socio-organizational perspectives,” *Information systems journal*, vol. 11, n^o. 2, p. 127–153, 2001.

- [38] K. L. Gwebu, J. Wang et M. Y. Hu, “Information security policy noncompliance : An integrative social influence model,” *Information Systems Journal*, vol. 30, n°. 2, p. 220–269, 2020.
- [39] M. T. Dlamini, J. H. Eloff et M. M. Eloff, “Information security : The moving target,” *Computers & security*, vol. 28, n°. 3-4, p. 189–198, 2009.
- [40] J. A. Addae, G. Simpson et G. O. A. Ampong, “Factors influencing information security policy compliance behavior,” dans *2019 International Conference on Cyber Security and Internet of Things (ICSIoT)*. IEEE, 2019, p. 43–47.
- [41] R. Palanisamy, A. A. Norman et M. L. M. Kiah, “Compliance with bring your own device security policies in organizations : A systematic literature review,” *Computers & Security*, vol. 98, p. 101998, 2020.
- [42] N. Ameen *et al.*, “Employees’ behavioural intention to smartphone security : A gender-based, cross-national study,” *Computers in Human Behavior*, vol. 104, p. 106184, 2020.
- [43] C. A. Agudelo-Serna *et al.*, “Strategies to mitigate knowledge leakage risk caused by the use of mobile devices : A preliminary study,” 2017.
- [44] N. S. Safa *et al.*, “Motivation and opportunity based model to reduce information security insider threats in organisations,” *Journal of information security and applications*, vol. 40, p. 247–257, 2018.
- [45] T. Sommestad, H. Karlzén et J. Hallberg, “The sufficiency of the theory of planned behavior for explaining information security policy compliance,” *Information & Computer Security*, 2015.
- [46] Hutto, Z. and Audet, C. (2021) Hype cycle for legal and compliance technologies. [En ligne]. Disponible : <https://www.gartner.com/interactive/hc/4003713?ref=solrAll&refval=318908894>
- [47] W. W. W. Consortium *et al.*, “Owl 2 web ontology language document overview,” 2012.
- [48] L. Jean-Baptiste, “Ontologies with python,” *Apress, Berkeley, CA*, 2021.
- [49] J.-B. Lamy, “Owlready : Ontology-oriented programming in python with automatic classification and high level constructs for biomedical ontologies,” *Artificial intelligence in medicine*, vol. 80, p. 11–28, 2017.
- [50] N. F. Noy, D. L. McGuinness *et al.*, “Ontology development 101 : A guide to creating your first ontology,” 2001.
- [51] F. Blackler, “Knowledge, knowledge work and organizations : An overview and interpretation,” *Organization studies*, vol. 16, n°. 6, p. 1021–1046, 1995.

- [52] M. Ivanović et Z. Budimac, “An overview of ontologies and data resources in medical domains,” *Expert Systems with Applications*, vol. 41, n^o. 11, p. 5158–5166, 2014.
- [53] M.-A. Léger, “Implementing information technologies with an ontology of cybersecurity professional skills,” Thèse de doctorat, 2022.
- [54] C. Bento, A. Cardoso et G. Dias, *Progress in Artificial Intelligence : 12th Portuguese Conference on Artificial Intelligence, EPIA 2005, Covilha, Portugal, December 5-8, 2005, Proceedings*. Springer, 2005, vol. 3808.
- [55] A. Oltramari *et al.*, “Building an ontology of cyber security.” dans *STIDS*. Citeseer, 2014, p. 54–61.
- [56] S.-W. Lee *et al.*, “Building problem domain ontology from security requirements in regulatory documents,” dans *Proceedings of the 2006 international workshop on Software engineering for secure systems*, 2006, p. 43–50.
- [57] K. Arbanas et M. Čubrić, “Ontology in information security,” *Journal of Information and Organizational Sciences*, vol. 39, n^o. 2, p. 107–136, 2015.
- [58] S. Fenz, S. Plieschnegger et H. Hobel, “Mapping information security standard iso 27002 to an ontological structure,” *Information & Computer Security*, vol. 24, n^o. 5, p. 452–473, 2016.
- [59] S. Ramanauskaitė *et al.*, “Security ontology for adaptive mapping of security standards,” *International Journal of Computers, Communications & Control (IJCCC)*, vol. 8, n^o. 6, p. 813–825, 2013.
- [60] F. de Franco Rosa et M. Jino, “A survey of security assessment ontologies,” dans *World Conference on Information Systems and Technologies*. Springer, 2017, p. 166–173.
- [61] L. Obrst, “Ontological architectures,” dans *Theory and applications of ontology : computer applications*. Springer, 2010, p. 27–66.
- [62] Z. Syed *et al.*, “Uco : A unified cybersecurity ontology,” dans *Workshops at the thirtieth AAAI conference on artificial intelligence*, 2016.
- [63] The MITRE Corporation . (2022) A knowledge graph of cybersecurity countermeasures. [En ligne]. Disponible : <https://d3fend.mitre.org/ontologies/d3fend.owl>
- [64] B. Andrei et O. Evgeny, “Framework for ontology-driven threat modelling of modern computer systems,” *International Journal of Open Information Technologies*, vol. 8, n^o. 2, p. 14–20, 2020.
- [65] A. Brazhuk, “Security patterns based approach to automatically select mitigations in ontology-driven threat modelling,” 2020.

- [66] D. Costa *et al.*, “An insider threat indicator ontology,” Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, Rapport technique CMU/SEI-2016-TR-007, 2016.
- [67] Enrico Fagnoni . (2011) It service management ontology, a standardized vocabulary for describing it services. [En ligne]. Disponible : <http://ontology.it/itsmo/v1/userguide/>
- [68] Marcus Bakker and Ruben Bouman . (2022) Detect tactics, techniques & combat threats. [En ligne]. Disponible : <https://github.com/rabobank-cdc/DeTTECT>
- [69] R. Ford *et al.*, “Automating financial regulatory compliance using ontology+ rules and sunflower,” dans *Proceedings of the 12th International Conference on Semantic Systems*, 2016, p. 113–120. [En ligne]. Disponible : <https://github.com/edmcouncil/fibo/>
- [70] Rinke Hoekstra . (2013) Lkif core. [En ligne]. Disponible : <https://github.com/RinkeHoekstra/lkif-core>
- [71] The MITRE Corporation . (2022) The mitre att&ck navigator. [En ligne]. Disponible : <https://mitre-attack.github.io/attack-navigator/>
- [72] N. Casellas *et al.*, “Ontological semantics for data privacy compliance : The neurona project,” dans *2010 AAAI Spring Symposium Series*, 2010.
- [73] ADAPT Centre . (2022) Gdpr compliance ontologies. [En ligne]. Disponible : <https://openscience.adaptcentre.ie/res/>
- [74] M. Palmirani *et al.*, “Legal ontology for modelling gdpr concepts and norms,” dans *Legal Knowledge and Information Systems*. IOS Press, 2018, p. 91–100.
- [75] M. Mockus. Methodology for building legal ontology. Accessed : 2022-02-12. [En ligne]. Disponible : <https://github.com/martynui/OGDL4M/wiki/Methodology>
- [76] K. P. Joshi *et al.*, “An ontology for a hipaa compliant cloud service,” dans *4th International IBM Cloud Academy Conference ICACON 2016*, 2016.
- [77] L. Alexander, “The ontology of consent,” *Analytic Philosophy*, p. 14–137, 2014.
- [78] S. Majumdar *et al.*, “Cloud security auditing : Major approaches and existing challenges,” dans *International Symposium on Foundations and Practice of Security*. Springer, 2018, p. 61–77.
- [79] R. Stevens *et al.*, “Compliance cautions : Investigating security issues associated with us digital-security standards.” dans *NDSS*, 2020.
- [80] Internal Revenue Service (IRS). (2023) Publication 1075 : Tax information security guidelines for federal, state and local agencies- safeguards for protecting federal tax returns and return information. [En ligne]. Disponible : <https://www.irs.gov/pub/irs-pdf/p1075.pdf>

- [81] North American Electric Reliability Corporation (NERC). (2023) Critical infrastructure protection. [En ligne]. Disponible : <https://www.nerc.com/comm/CIPC/Pages/default.aspx>
- [82] N. A. A. Othman, A. A. Norman et M. L. M. Kiah, “Systematic literature review of security control assessment challenges,” dans *2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE, 2022, p. 31–36.
- [83] D. IMERI, “The standardization vs. customization debate continues for pci dss compliant products,” 2015.
- [84] A. Buecker *et al.*, *IT Security Compliance Management Design Guide with IBM Tivoli Security Information and Event Manager*. IBM Redbooks, 2010.
- [85] M. Jans, M. G. Alles et M. A. Vasarhelyi, “A field study on the use of process mining of event logs as an analytical procedure in auditing,” *The Accounting Review*, vol. 89, n^o. 5, p. 1751–1773, 2014.
- [86] S. Jakoubi *et al.*, “A survey of scientific approaches considering the integration of security and risk aspects into business process management,” dans *2009 20th International Workshop on Database and Expert Systems Application*. IEEE, 2009, p. 127–132.
- [87] J. S. Gill *et al.*, “Active policy enforcement,” juill. 10 2018, uS Patent 10,019,677.
- [88] M. Wurzenberger *et al.*, “Aecid : A self-learning anomaly detection approach based on light-weight log parser models.” dans *ICISSP*, 2018, p. 386–397.
- [89] ———, “Aecid-pg : A tree-based log parser generator to enable log analysis,” dans *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 2019, p. 7–12.
- [90] M. Landauer *et al.*, “A framework for cyber threat intelligence extraction from raw log data,” dans *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, p. 3200–3209.
- [91] F. Skopik, M. Wurzenberger et M. Landauer, *Smart Log Data Analytics : Techniques for Advanced Security Analysis*. Springer, 2021.
- [92] M. Landauer *et al.*, “Have it your way : Generating customized log datasets with a model-driven simulation testbed,” *IEEE Transactions on Reliability*, vol. 70, n^o. 1, p. 402–415, 2020.
- [93] M. Wurzenberger *et al.*, “Creating character-based templates for log data to enable security event classification,” dans *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, 2020, p. 141–152.

- [94] Austrian Institute of Technology (AIT). (2023) Research papers and industrial patents that support aminer and aecid tools. [En ligne]. Disponible : <https://aecid.ait.ac.at/further-information/>
- [95] N. McCurdy, J. Dykes et M. Meyer, “Action design research and visualization design,” dans *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 2016, p. 10–18.
- [96] M. T. Mullarkey et A. R. Hevner, “An elaborated action design research process model,” *European Journal of Information Systems*, vol. 28, n^o. 1, p. 6–20, 2019.
- [97] A. R. Hevner *et al.*, “Design science in information systems research,” *MIS quarterly*, p. 75–105, 2004.
- [98] J. Kukulies, B. Falk et R. H. Schmitt, “Development of optimized test planning procedures for stabilizing ramp-up processes by means of design science research,” *Procedia CIRP*, vol. 51, p. 93–98, 2016.
- [99] F. Niederman et S. T. March, “Design science and the accumulation of knowledge in the information systems discipline,” *ACM Transactions on Management Information Systems (TMIS)*, vol. 3, n^o. 1, p. 1–15, 2012.
- [100] W. J. Keijzer-Broers et M. de Reuver, “Action design research for social innovation : Lessons from designing a health and wellbeing platform.” dans *ICIS*, 2016.
- [101] M. K. Sein *et al.*, “Action design research,” *MIS quarterly*, p. 37–56, 2011.
- [102] J. E. v. Aken, “Management research based on the paradigm of the design sciences : the quest for field-tested and grounded technological rules,” *Journal of management studies*, vol. 41, n^o. 2, p. 219–246, 2004.
- [103] M. Vianden *et al.*, “History and lessons learnt from a metrics program at a cmmi level 3 company,” dans *2013 20th Asia-Pacific Software Engineering Conference (APSEC)*, vol. 2. IEEE, 2013, p. 73–78.
- [104] G. Ridley, J. Young et P. Carroll, “Cobit and its utilization : A framework from the literature,” dans *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. IEEE, 2004, p. 8–pp.
- [105] Wazuh. Wazuh github repo. Accessed : 2023-02-23. [En ligne]. Disponible : <https://github.com/wazuh/wazuh/tree/master/ruleset>
- [106] ——. Wazuh ruleset documentation. Accessed : 2023-02-23. [En ligne]. Disponible : <https://documentation.wazuh.com/current/user-manual/ruleset/index.html>
- [107] S. Stanković, S. Gajin et R. Petrović, “A review of wazuh tool capabilities for detecting attacks based on log analysis.”

- [108] Syeda Wardah Batool . (2021) How to use linux logger command. [En ligne]. Disponible : <https://linuxhint.com/use-logger-command-linux/>
- [109] Ubuntu Manual Pages . (2019) Logger documentation. [En ligne]. Disponible : <https://manpages.ubuntu.com/manpages/bionic/man1/logger.1.html>
- [110] Dataguard. (2022) Annex a.9 : Access control (iso27001). [En ligne]. Disponible : <https://www.dataguard.co.uk/blog/iso-27001-annex-a.9-access-control/>
- [111] A. Security. Cloudsploit. Accessed : 2022-02-13. [En ligne]. Disponible : <https://github.com/aquasecurity/cloudsploit>
- [112] N. G. Plc. Scoutsuite. Accessed : 2022-02-13. [En ligne]. Disponible : <https://github.com/nccgroup/ScoutSuite>
- [113] Prowler. Prowler. Accessed : 2022-02-13. [En ligne]. Disponible : <https://github.com/prowler-cloud/prowler>
- [114] M. D. Wilkinson *et al.*, “The fair guiding principles for scientific data management and stewardship,” *Scientific data*, vol. 3, n^o. 1, p. 1–9, 2016.
- [115] C. Brewster *et al.*, “Ontology-based access control for fair data,” *Data Intelligence*, vol. 2, n^o. 1-2, p. 66–77, 2020.
- [116] A. Mohammad *et al.*, “Ontology-based access control model for semantic web service,” *Journal of Information and Computing Science*, vol. 6, n^o. 3, p. 177–194, 2011.
- [117] W3 Community. (2023) Webaccesscontrol. [En ligne]. Disponible : <https://www.w3.org/wiki/WebAccessControl>
- [118] Linkeddata. (2023) Basic access control ontology. [En ligne]. Disponible : <https://lov.linkeddata.es/dataset/lov/vocabs/acl>
- [119] M. Buffa et C. Faron-Zucker, “Ontology-based access rights management,” dans *Advances in Knowledge Discovery and Management : Volume 2*. Springer, 2012, p. 49–61.
- [120] E. Kalemi et E. Martiri, “Foaf-academic ontology : a vocabulary for the academic community,” dans *2011 Third International Conference on Intelligent Networking and Collaborative Systems*. IEEE, 2011, p. 440–445.
- [121] L. Elluri, A. Nagar et K. P. Joshi, “An integrated knowledge graph to automate gdpr and pci dss compliance,” dans *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, p. 1266–1271.
- [122] B. McBride, “The resource description framework (rdf) and its vocabulary description language rdfs,” *Handbook on ontologies*, p. 51–65, 2004.
- [123] M. Horridge *et al.*, “A practical guide to building owl ontologies using protégé 4 and co-ode tools edition1. 2,” *The university of Manchester*, vol. 107, 2009.

- [124] R. Team. Rdfliib python module. Accessed : 2022-02-13. [En ligne]. Disponible : <https://rdfliib.readthedocs.io/en/stable/>
- [125] Michele Pasin. (2022) Ontospy : Python library and command-line interface for inspecting and visualizing rdf models. [En ligne]. Disponible : <https://pypi.org/project/ontospy/>
- [126] R. Vaarandi, “A data clustering algorithm for mining patterns from event logs,” dans *Proceedings of the 3rd IEEE Workshop on IP Operations & Management (IPOM 2003)*(IEEE Cat. No. 03EX764). Ieee, 2003, p. 119–126.
- [127] C. Zhuge et R. Vaarandi, “Efficient event log mining with logcluster,” dans *2017 ieee 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids)*. IEEE, 2017, p. 261–266.
- [128] M. Wurzenberger *et al.*, “Applying high-performance bioinformatics tools for outlier detection in log data,” dans *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*. IEEE, 2017, p. 1–8.
- [129] V. R. Basili, “Applying the goal/question/metric paradigm in the experience factory,” *Software quality assurance and measurement : A worldwide perspective*, vol. 7, n^o. 4, p. 21–44, 1993.
- [130] V. R. B. G. Caldiera et H. D. Rombach, “The goal question metric approach,” *Encyclopedia of software engineering*, p. 528–532, 1994.
- [131] R. G. Utomo, G. Wills et R. Walters, “Developing an instrument to measure information assurance implementation for egovernment using goal question metric approach,” *Kinetik : Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 2021.
- [132] N. Bhensook et T. Senivongse, “An assessment of security requirements compliance of cloud providers,” dans *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings*. IEEE, 2012, p. 520–525.
- [133] M. Corpuz, “Enterprise information security policy assessment-an extended framework for metrics development utilising the goal-question-metric approach,” *IS Institute, Queensland University of Technology, Brisbane, Queensland, Australia,(2011)*, 2011.
- [134] Living Security Team . (2021) 6 metrics to track in your cybersecurity awareness training campaign. [En ligne]. Disponible : <https://www.livingsecurity.com/blog/metrics-to-track-in-your-cybersecurity-awareness-training-campaign>

- [135] J. P. de Brito Gonçalves *et al.*, “A quality of service compliance system empowered by smart contracts and oracles,” dans *2020 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2020, p. 532–538.
- [136] T. C. C. Blog. 5 compliance metrics you should use for effective product compliance management. Accessed : 2022-02-10. [En ligne]. Disponible : <https://blog.clevercompliance.io/product-compliance/5-compliance-metrics-you-should-use-for-effective-product-compliance-management/>
- [137] M. Stream. Reporting compliance metrics that matter to the management and the board. Accessed : 2022-02-11. [En ligne]. Disponible : <https://www.metricstream.com/insights/reporting-compliance-metrics.htm>
- [138] H.-P. Lam, M. Hashmi et A. Kumar, “Towards a formal framework for partial compliance of business processes,” dans *AI Approaches to the Complexity of Legal Systems XI-XII : AICOL International Workshops 2018 and 2020 : AICOL-XI@ JURIX 2018, AICOL-XII@ JURIX 2020, XAILA@ JURIX 2020, Revised Selected Papers*. Springer, 2021, p. 90–105.
- [139] S. Sadiq, G. Governatori et K. Namiri, “Modeling control objectives for business process compliance,” dans *Business Process Management : 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007. Proceedings 5*. Springer, 2007, p. 149–164.
- [140] E. Morrison, A. Ghose et G. Koliadis, “Dealing with imprecise compliance requirements,” dans *2009 13th Enterprise Distributed Object Computing Conference Workshops*. IEEE, 2009, p. 6–14.
- [141] S. Dännart, F. M. Constante et K. Beckers, “An assessment model for continuous security compliance in large scale agile environments : exploratory paper,” dans *Advanced Information Systems Engineering : 31st International Conference, CAiSE 2019, Rome, Italy, June 3–7, 2019, Proceedings 31*. Springer, 2019, p. 529–544.
- [142] A. E. Martinez Suñé et C. G. Lopez Pombo, “Quality of service ranking by quantifying partial compliance of requirements,” dans *Coordination Models and Languages : 22nd IFIP WG 6.1 International Conference, COORDINATION 2020, Held as Part of the 15th International Federated Conference on Distributed Computing Techniques, DisCoTec 2020, Valletta, Malta, June 15–19, 2020, Proceedings*. Springer, 2020, p. 181–189.
- [143] Gartner Care . (2022) Gartner care report. [En ligne]. Disponible : <https://www.gartner.com/document/4003605?ref=solrAll&refval=336347991>

- [144] S. Rinderle-Ma et K. Winter, “Predictive compliance monitoring in process-aware information systems : State of the art, functionalities, research directions,” *arXiv preprint arXiv :2205.05446*, 2022.
- [145] J. Kim et M. Comuzzi, “Stability metrics for enhancing the evaluation of outcome-based business process predictive monitoring,” *IEEE Access*, vol. 9, p. 133 461–133 471, 2021.
- [146] C. Klinkmüller, N. R. van Beest et I. Weber, “Towards reliable predictive process monitoring,” dans *Information Systems in the Big Data Era : CAiSE Forum 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings 30*. Springer, 2018, p. 163–181.
- [147] M. De Leoni, W. M. van der Aalst et M. Dees, “A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs,” *Information Systems*, vol. 56, p. 235–257, 2016.
- [148] C. Sillaber, A. Mussmann et R. Breu, “Experience : Data and information quality challenges in governance, risk, and compliance management,” *Journal of Data and Information Quality (JDIQ)*, vol. 11, n°. 2, p. 1–14, 2019.
- [149] J. Frankland, “It security metrics : implementation and standards compliance,” *Network Security*, vol. 2008, n°. 6, p. 6–9, 2008.
- [150] D. Coffman, B. Agrawal et F. Schaffa, “Towards optimal risk-aware security compliance of a large it system,” dans *Service-Oriented Computing : 11th International Conference, ICSOC 2013, Berlin, Germany, December 2-5, 2013, Proceedings 11*. Springer, 2013, p. 639–651.
- [151] Andrej Volchkov . (2020) Key performance indicators for security governance, part 2 : Security reporting for senior management. [En ligne]. Disponible : <https://www.isaca.org/resources/isaca-journal/issues/2020/volume-6/key-performance-indicators-for-security-governance-part-2>
- [152] SecurityScoreCard . (2019) 20 cybersecurity metrics & kpis to track in 2022. [En ligne]. Disponible : <https://securityscorecard.com/blog/9-cybersecurity-metrics-kpis-to-track>
- [153] Cipher. (2021) 10 cybersecurity metrics you should be monitoring. [En ligne]. Disponible : <https://cipher.com/blog/10-cybersecurity-metrics-you-should-be-monitoring/>
- [154] Cyphere. (2021) Top cybersecurity metrics (kpi) to track. [En ligne]. Disponible : <https://thecyphere.com/blog/security-kpi/>
- [155] Jose Samuel. (2019) Cyber security — key performance indicators. [En ligne]. Disponible : <https://infosecwriteups.com/cyber-security-key-performance-indicators-30971d04672>

- [156] Andrew Froehlich. (2021) 7 key cybersecurity metrics for the board and how to present them. [En ligne]. Disponible : <https://www.techtarget.com/searchsecurity/tip/7-key-cybersecurity-metrics-for-the-board-and-how-to-present-them>
- [157] Ellen Zhang, DigitalGuardian. (2017) Security and analytics experts share the most important cybersecurity metrics and kpis. [En ligne]. Disponible : <https://digitalguardian.com/blog/what-are-the-most-important-cybersecurity-metrics-kpis>
- [158] Kaitlyn Archibald. (2022) How are you measuring infosec kris and cybersecurity metrics? [En ligne]. Disponible : <https://www.onetrust.com/blog/how-are-you-measuring-infosec-kris-and-cybersecurity-metrics/>
- [159] V. Anu, “Information security governance metrics : a survey and taxonomy,” *Information Security Journal : A Global Perspective*, vol. 31, n°. 4, p. 466–478, 2022.
- [160] R. P. Mganga, M. Charles *et al.*, “Enhancing information security in cloud computing services using sla based metrics,” 2011.
- [161] T. D. Breaux *et al.*, “It compliance : aligning legal and product requirements,” *IT professional*, vol. 11, n°. 5, p. 54–58, 2009.
- [162] M. R. I. Nekvi, I. Noorwali et N. H. Madhavji, “Deriving metrics for estimating the effort needed in requirements compliance work,” dans *Requirements Engineering : Foundation for Software Quality : 22nd International Working Conference, REFSQ 2016, Gothenburg, Sweden, March 14-17, 2016, Proceedings 22*. Springer, 2016, p. 135–141.
- [163] U. Ozeer, “ φ comp : An architecture for monitoring and enforcing security compliance in sensitive health data environment,” dans *2021 IEEE 18th International Conference on Software Architecture Companion (ICSA-C)*. IEEE, 2021, p. 70–77.
- [164] C. Abid, M. Kessentini et H. Wang, “Early quality of service prediction via interface-level metrics, code-level metrics, and antipatterns,” 2020.
- [165] L. N. Smith, “A disciplined approach to neural network hyper-parameters : Part 1–learning rate, batch size, momentum, and weight decay,” *arXiv preprint arXiv :1803.09820*, 2018.
- [166] M. Y. Park et T. Hastie, “L1-regularization path algorithm for generalized linear models,” *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 69, n°. 4, p. 659–677, 2007.
- [167] K. Weiss, T. M. Khoshgoftaar et D. Wang, “A survey of transfer learning,” *Journal of Big data*, vol. 3, n°. 1, p. 1–40, 2016.
- [168] Mavs, *ATOM : A Python package for fast exploration of machine learning pipelines*, 2019, aTOM version 2.0.3. [En ligne]. Disponible : <https://tvdbboom.github.io/ATOM/>

- [169] P. Lai *et al.*, “Ontology-based interpretable machine learning for textual data,” dans *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, p. 1–10.
- [170] F. Pedregosa *et al.*, “Scikit-learn : Machine learning in python,” *the Journal of machine Learning research*, vol. 12, p. 2825–2830, 2011.
- [171] T. Developers, “Tensorflow,” *Zenodo*, 2022.
- [172] A. Paszke *et al.*, “Pytorch : An imperative style, high-performance deep learning library,” *Advances in neural information processing systems*, vol. 32, 2019.
- [173] S. Qaiser et R. Ali, “Text mining : use of tf-idf to examine the relevance of words to documents,” *International Journal of Computer Applications*, vol. 181, n^o. 1, p. 25–29, 2018.
- [174] M. Abbas *et al.*, “Multinomial naive bayes classification model for sentiment analysis,” *IJCSNS Int. J. Comput. Sci. Netw. Secur*, vol. 19, n^o. 3, p. 62, 2019.
- [175] S. Wang, L. Jiang et C. Li, “Adapting naive bayes tree for text classification,” *Knowledge and Information Systems*, vol. 44, p. 77–89, 2015.
- [176] A. Anagaw et Y.-L. Chang, “A new complement naïve bayesian approach for biomedical data classification,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, p. 3889–3897, 2019.
- [177] S. L. Cessie et J. V. Houwelingen, “Ridge estimators in logistic regression,” *Journal of the Royal Statistical Society Series C : Applied Statistics*, vol. 41, n^o. 1, p. 191–201, 1992.
- [178] D. W. Hosmer Jr, S. Lemeshow et R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.
- [179] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, n^o. 2, p. 1883, 2009.
- [180] M. Muja et D. G. Lowe, “Scalable nearest neighbor algorithms for high dimensional data,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, n^o. 11, p. 2227–2240, 2014.
- [181] S. R. Safavian et D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE transactions on systems, man, and cybernetics*, vol. 21, n^o. 3, p. 660–674, 1991.
- [182] C. D. Sutton, “Classification and regression trees, bagging, and boosting,” *Handbook of statistics*, vol. 24, p. 303–329, 2005.
- [183] M. W. Ahmad, J. Reynolds et Y. Rezgui, “Predictive modelling for solar thermal energy systems : A comparison of support vector regression, random forest, extra trees and regression trees,” *Journal of cleaner production*, vol. 203, p. 810–821, 2018.

- [184] L. Breiman, “Random forests,” *Machine learning*, vol. 45, p. 5–32, 2001.
- [185] X. Li, L. Wang et E. Sung, “Adaboost with svm-based component classifiers,” *Engineering Applications of Artificial Intelligence*, vol. 21, n^o. 5, p. 785–795, 2008.
- [186] A. Natekin et A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in neuro-robotics*, vol. 7, p. 21, 2013.
- [187] T. Chen et C. Guestrin, “Xgboost : A scalable tree boosting system,” dans *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, p. 785–794.
- [188] G. Ke *et al.*, “Lightgbm : A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems*, vol. 30, 2017.
- [189] L. Prokhorenkova *et al.*, “Catboost : unbiased boosting with categorical features,” *Advances in neural information processing systems*, vol. 31, 2018.
- [190] T. Joachims, “Training linear svms in linear time,” dans *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, p. 217–226.
- [191] M. Murty *et al.*, “Kernel-based svm,” *Support vector machines and perceptrons : learning, optimization, classification, and application to social networks*, p. 57–67, 2016.
- [192] K. Crammer *et al.*, “Online passive aggressive algorithms,” 2006.
- [193] L. Bottou, “Stochastic gradient descent tricks,” *Neural Networks : Tricks of the Trade : Second Edition*, p. 421–436, 2012.
- [194] J. Davis et M. Goadrich, “The relationship between precision-recall and roc curves,” dans *Proceedings of the 23rd international conference on Machine learning*, 2006, p. 233–240.
- [195] D. Chicco et G. Jurman, “The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation,” *BMC genomics*, vol. 21, p. 1–13, 2020.
- [196] A. Moffat et J. Zobel, “Rank-biased precision for measurement of retrieval effectiveness,” *ACM Transactions on Information Systems (TOIS)*, vol. 27, n^o. 1, p. 1–27, 2008.
- [197] S. Niwattanakul *et al.*, “Using of jaccard coefficient for keywords similarity,” dans *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1, n^o. 6, 2013, p. 380–384.
- [198] S. Visa *et al.*, “Confusion matrix-based feature selection.” *Maics*, vol. 710, n^o. 1, p. 120–127, 2011.

- [199] F. Lécué et J. Wu, “Semantic explanations of predictions,” *arXiv preprint arXiv :1805.10587*, 2018.
- [200] R. Confalonieri *et al.*, “An ontology-based approach to explaining artificial neural networks,” *arXiv preprint arXiv :1906.08362*, 2019.
- [201] A. Ponomarev et A. Agafonov, “Ontology concept extraction algorithm for deep neural networks,” dans *2022 32nd Conference of Open Innovations Association (FRUCT)*. IEEE, 2022, p. 221–226.
- [202] A. Yates *et al.*, “Texrunner : open information extraction on the web,” dans *Proceedings of Human Language Technologies : The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2007, p. 25–26.
- [203] M. Schmitz *et al.*, “Open language learning for information extraction,” dans *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, 2012, p. 523–534.
- [204] O. Etzioni *et al.*, “Open language learning for information extraction,” oct. 2 2014, uS Patent App. 14/083,342.
- [205] E. Dabbas, *Interactive Dashboards and Data Apps with Plotly and Dash : Harness the power of a fully fledged frontend web framework in Python—no JavaScript required*. Packt Publishing Ltd, 2021.
- [206] Python Community. (2023) Plotly dash. [En ligne]. Disponible : <https://dash.plotly.com/>
- [207] Professeur Michel Gagnon. (2023) Cours : Inf8410 - ontologie et web sémantique. [En ligne]. Disponible : <https://www.polymtl.ca/programmes/cours/ontologie-et-web-semantique>
- [208] M. Poveda-Villalón, M. C. Suárez-Figueroa et A. Gómez-Pérez, “Validating ontologies with oops!” dans *Knowledge Engineering and Knowledge Management : 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings 18*. Springer, 2012, p. 267–281.
- [209] E. Sirin et B. Parsia, “Sparql-dl : Sparql query for owl-dl.” dans *OWLED*, vol. 258, 2007.
- [210] Jayakumar Sundaram. (2017) The benefits of the statement of applicability in isms projects. [En ligne]. Disponible : <https://www.isaca.org/resources/isaca-journal/issues/2017/volume-3/the-benefits-of-the-statement-of-applicability-in-isms-projects>

- [211] M. Dorato et M. Esfeld, “Grw as an ontology of dispositions,” *Studies in History and Philosophy of Science Part B : Studies in History and Philosophy of Modern Physics*, vol. 41, n^o. 1, p. 41–49, 2010.
- [212] S. Howell, T. Beach et Y. Rezgui, “Robust requirements gathering for ontologies in smart water systems,” *Requirements Engineering*, vol. 26, n^o. 1, p. 97–114, 2021.
- [213] J. Potoniec *et al.*, “Dataset of ontology competency questions to sparql-owl queries translations,” *Data in brief*, vol. 29, p. 105098, 2020.
- [214] M. Grüninger et M. S. Fox, “The role of competency questions in enterprise engineering,” *Benchmarking—Theory and practice*, p. 22–31, 1995.
- [215] R. K. Yin, “Validity and generalization in future case study evaluations,” *Evaluation*, vol. 19, n^o. 3, p. 321–332, 2013.
- [216] A. Richter et C. Koch, “Integration, differentiation and ambiguity in safety cultures,” *Safety science*, vol. 42, n^o. 8, p. 703–722, 2004.
- [217] R. Whitemore, S. K. Chase et C. L. Mandle, “Validity in qualitative research,” *Qualitative health research*, vol. 11, n^o. 4, p. 522–537, 2001.
- [218] G. Governatori, “Representing business contracts in ruleml,” *International Journal of Cooperative Information Systems*, vol. 14, n^o. 02n03, p. 181–216, 2005.
- [219] Assemblée Nationale du Québec. (2023) Loi modernisant des dispositions législatives en matière de protection des renseignements personnels, lq 2021, c 25. [En ligne]. Disponible : <https://www.canlii.org/fr/qc/legis/loisa/lq-2021-c-25/derniere/lq-2021-c-25.html>
- [220] K. Stouffer *et al.*, “Guide to operational technology (ot) security,” National Institute of Standards and Technology, Rapport technique, 2022.
- [221] P. N. Otto et A. I. Antón, “Addressing legal requirements in requirements engineering,” dans *15th IEEE international requirements engineering conference (RE 2007)*. IEEE, 2007, p. 5–14.
- [222] T. D. Breaux, M. W. Vail et A. I. Anton, “Towards regulatory compliance : Extracting rights and obligations to align requirements with regulations,” dans *14th IEEE International Requirements Engineering Conference (RE’06)*. IEEE, 2006, p. 49–58.
- [223] M. J. May, C. A. Gunter et I. Lee, “Privacy apis : Access control techniques to analyze and verify legal privacy policies,” dans *19th IEEE Computer Security Foundations Workshop (CSFW’06)*. IEEE, 2006, p. 13–pp.
- [224] F. Massacci, M. Prest et N. Zannone, “Using a security requirements engineering methodology in practice : the compliance with the italian data protection legislation,” *Computer Standards & Interfaces*, vol. 27, n^o. 5, p. 445–455, 2005.

- [225] J. C. Maxwell et A. I. Antón, “Developing production rule models to aid in acquiring requirements from legal texts,” dans *2009 17th IEEE International requirements engineering conference*. IEEE, 2009, p. 101–110.
- [226] J. Cleland-Huang *et al.*, “A machine learning approach for tracing regulatory codes to product specific requirements,” dans *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering- Volume 1*, 2010, p. 155–164.
- [227] J. Guo, J. Cleland-Huang et B. Berenbach, “Foundations for an expert system in domain-specific traceability,” dans *2013 21st IEEE International Requirements Engineering Conference (RE)*. IEEE, 2013, p. 42–51.
- [228] J. C. Maxwell et A. I. Antón, “Checking existing requirements for compliance with law using a production rule model,” dans *2009 second international workshop on requirements engineering and law*. IEEE, 2009, p. 1–6.
- [229] T. D. Breaux, *Legal requirements acquisition for the specification of legally compliant information systems*. North Carolina State University, 2009.
- [230] J.-M. Hufflen, “{BibTeX},{mlBibTeX} and bibliography styles,” *Biuletyn {GUST}*, p. 76–80, 2006.
- [231] L. C. Team. Litmaps. Accessed : 2022-02-13. [En ligne]. Disponible : <https://www.litmaps.co/>
- [232] S. Keshav, “How to read a paper,” *ACM SIGCOMM Computer Communication Review*, vol. 37, n°. 3, p. 83–84, 2007.
- [233] K. Mueen Ahmed et B. E. A. Dhubaib, “Zotero : A bibliographic assistant to researcher,” *Journal of Pharmacology and Pharmacotherapeutics*, vol. 2, n°. 4, p. 304–305, 2011.
- [234] P. M. Library. Guide on literature review. Accessed : 2022-02-13. [En ligne]. Disponible : https://libguides.biblio.polymtl.ca/literature_review
- [235] ——. Finding scientific articles. Accessed : 2022-02-13. [En ligne]. Disponible : https://libguides.biblio.polymtl.ca/finding_scientific_articles
- [236] M.-A. Léger. Introduction à la recherche de littérature. Accessed : 2022-02-13. [En ligne]. Disponible : https://www.youtube.com/watch?v=5d90hhLNcu8&ab_channel=Marc-AndreLeger
- [237] F. d. Franco Rosa et M. Jino, “A survey of security assessment ontologies,” dans *World Conference on Information Systems and Technologies*. Springer, 2017, p. 166–173.
- [238] A. Harzing. (2007) Publish or perish software. Accessed : 2022-02-13. [En ligne]. Disponible : <https://harzing.com/resources/publish-or-perish>

- [239] B. Kitchenham *et al.*, “Systematic literature reviews in software engineering—a systematic literature review,” *Information and software technology*, vol. 51, n^o. 1, p. 7–15, 2009.

ANNEXE 1: STRATÉGIE UTILISÉE POUR LA REVUE DE LITTÉRATURE

Nous avons entrepris une revue de la littérature avec une stratégie bien définie afin de nous assurer de la pertinence des articles pour notre étude. Nous avons pris en compte divers facteurs, notamment la pertinence du sujet des articles et la pertinence des personnes et des sites. Nous avons également considéré la pertinence du problème et de la question, en nous assurant que la littérature examinait le même problème de recherche et abordait la même question que notre étude proposée. Nous avons aussi tenu compte de l'accessibilité de la littérature, que ce soit à la bibliothèque de Polytechnique Montréal ou en ligne.

Afin de maintenir l'élan, nous avons créé un calendrier pour aider à la revue de la littérature. Le calendrier contient des dates, des jalons et des incitations pour aider à maintenir l'examen de la littérature sur la bonne voie. Le calendrier est basé sur les trois mêmes domaines indiqués dans le paragraphe de l'ouverture. Chaque domaine comprend au moins trois documents de recherche empirique liés à ce domaine.

Nous avons effectué une revue de la littérature en utilisant des sources primaires et en évitant les sources secondaires. Nous avons également créé une carte récapitulative de chaque article, triant les cartes dans un ordre logique qui raconte « l'histoire » de la recherche. Tout au long du processus, nous n'arrêtons pas de se demander : « Comment ce document de recherche est-il lié au problème de recherche ? » pour s'assurer que chaque article était approprié pour la revue de la littérature.

La revue de la littérature a été réalisée en utilisant les bases de données disponibles sur le site web de la bibliothèque de Polytechnique Montréal et sur Internet. Ces bases de données répertorient et identifient les articles scientifiques qui ont été publiés en anglais au cours des dix dernières années. Des mots-clés tels que «IT» et «Compliance» ont été utilisés dans les recherches. Les articles les plus fréquemment cités ont été sélectionnés et exportés au format Bibtex [230]. Ensuite, les fichiers Bibtex ont été importés dans LitMap [231], un programme de cartographie de référence. Sur la base de ce qui s'y trouvait, l'outil LitMap a ensuite identifié d'autres articles potentiellement liés. Cette application génère les choix les plus adaptés à l'aide de bases de données graphiques(Cf. Figure .1).

Une fois que nous avons les propositions de LitMap [231], nous suivons la stratégie suivante :

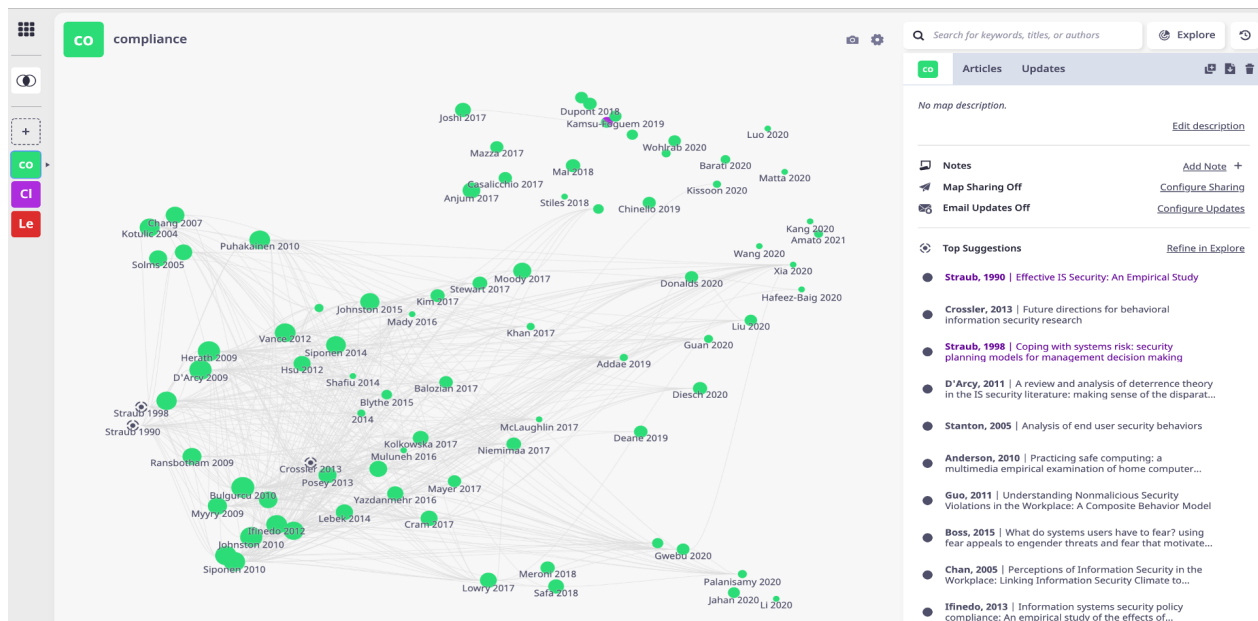


FIGURE .1 Carte de la littérature avec LitMap.

La figure montre une carte de littérature générée par LitMap, un programme en ligne de cartographie de références de recherche. Cette carte présente les références utilisées et analysées en relation avec le mot clé “Compliance”. Les différentes références sont représentées par des points sur la carte et sont reliées entre elles pour montrer les liens entre les différentes études.

1. Identification des publications qui pourraient être incluses dans le projet
2. Lecture des résumés d’articles fournis par les auteurs ou les éditeurs pour une validation préliminaire de la pertinence de l’article.
3. Si l’article semble pertinent après avoir lu le résumé et appliqué la stratégie de lecture en trois étapes recommandée dans [232], nous suivons les sous-étapes suivantes :
 - (a) Insertion de l’article dans le logiciel de gestion de références Zotero [233].
 - (b) Recherche d’une version PDF de l’article.
 - (c) Enregistrement de l’article au format PDF dans un dossier Dropbox partagé entre les membres de l’équipe de recherche.
 - (d) Première lecture de l’article dans son intégralité.
 - (e) Découverte d’éléments d’intérêt dans l’article.
4. Dans Zotero, les dossiers contenant les éléments d’intérêt sont produits et suivis. Dans le premier cas, les éléments d’information intéressants sont notés dans des notes jointes aux articles de référence correspondants.

5. Les concepts issus de l'évaluation initiale sont utilisés pour développer et alimenter une carte conceptuelle.
6. La revue de littérature initiale est entamée et améliorée avec les éléments indiqués.

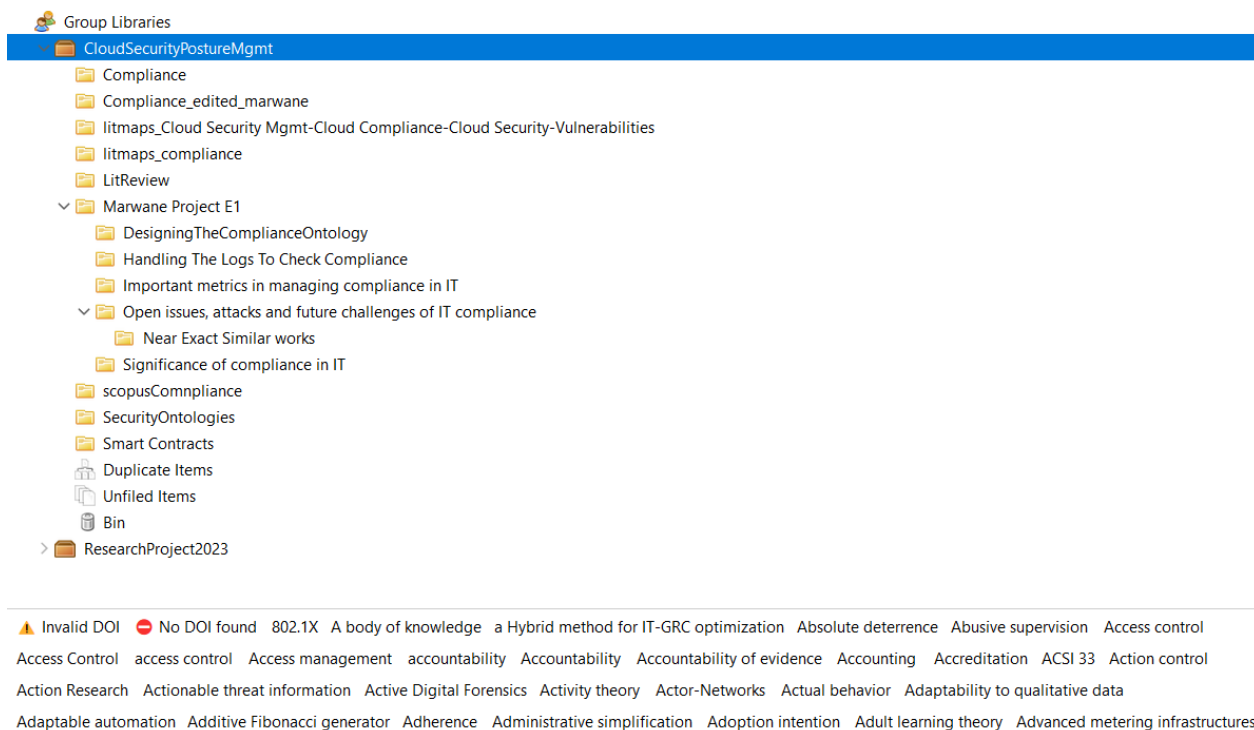


FIGURE .2 Sources et dossiers bibliographiques dans Zotero.

Cette figure montre comment les références bibliographiques ont été organisées en dossiers pour faciliter leur traitement. Par exemple, nous pouvons voir des dossiers tels que "Signification de la conformité dans les TI" et "Métriques importantes pour la gestion de la conformité dans les TI". Cela permet de regrouper les références par sujet et de les retrouver et traiter plus facilement.

Pour effectuer la revue de la littérature, les directives fournies par [234], [235], et [236] ont été soigneusement suivies. En outre, une approche similaire à [237] a été utilisée dans cette étude. Sur la base de l'objectif principal et des sous-objectifs de la recherche, les cinq besoins d'information suivants peuvent être identifiés :

1. Signification de la conformité informatique
2. Paramètres importants dans la gestion de la conformité des TI
3. Traitement des journaux pour vérifier la conformité des TI
4. Conception de l'ontologie de la conformité informatique
5. Questions ouvertes, attaques et défis futurs de la conformité informatique

Le lecteur est invité à consulter les figures .3 et .4 pour voir un exemple de la façon de répondre et de traiter un besoin d'information spécifique. IEEE, Google Scholar, Scopus, Engineering Village, ACM, ScienceDirect et Elsevier figuraient parmi les bases de données scientifiques les plus fiables au cours de cette revue de la littérature. Le logiciel Publish or Perish [238] a été utilisé pour obtenir un accès unifié à de nombreuses bases de données de recherche bien connues et pour classer les articles en fonction du nombre de citations et de l'index h des chercheurs.

Formulation of information needs:

What is the significance of compliance in IT? The types of compliance in IT?

Complian*	IT	Type	Significance	Information system	norms	Cyber security
Conform* Obedience Agree*	Information technology Tech business High-tech high technology	Kind Categor* Group* Classif* Variet* Sort Model Division Famili* Subdivision way	Sense Meaning Definition Essence Explanation Explication Interpretation Nuance Understanding Value Gist Core	IS corporate information system	Reference auditing procedure Standards Common patterns Rules Regulation Convention Principles Laws Specification Spec Security Standards and Certificates certificat* Compliance requirements	Cybersecurity data security information security"

FIGURE .3 Tableau représentant les chaînes de recherche utilisées pour un besoin d'information spécifique.

Le tableau en question décrit les chaînes de recherche utilisées pour répondre à un besoin d'information spécifique concernant la signification de la conformité dans les TI et les types de conformité dans ce domaine. Il montre comment les mots clés importants et les synonymes pertinents ont été identifiés pour traiter ce besoin. Ces informations ont ensuite été utilisées pour créer des commandes de recherche d'experts dans Engineering Village comme c'est indiqué dans la Figure .4.

Application de l'approche Systematic Literature Review (SLR)

L'approche Systematic Literature Review (SLR) [239] a été utilisée pour examiner et synthétiser les recherches existantes sur les métriques de conformité informatique. Nous avons suivi les étapes suivantes pour mener une SLR approfondie :

(Complian* OR compliance requirement* OR Conform* OR Obedien* OR Agree*) WN KY AND (IT OR Information technology OR tech business OR technolog* business OR high-tech OR high tech or high technolog* OR information system* OR IS OR corporate information system*) WN KY AND (Type OR Kind OR Categor* OR Group* OR Classif* OR Variet* OR Sort OR Model OR Division OR Famil* OR Subdivision OR way) WN KY AND (significan* OR Sense OR Mean* OR Definition* OR Essence* OR Explanation* OR Explication* OR Interpretation* OR Nuance* OR Understand* OR Value* OR Gist* OR Core*) WN KY AND (norm* OR Reference* OR audit* procedure* OR Standard* OR Common pattern* OR pattern* OR Rule* OR Regulation* OR Convention* OR Principle* OR Law* OR specification* OR spec OR Secur* Standard* OR certificat* OR Complian* requirement*) WN KY AND (Cybersecurity OR cyber security OR data security OR information security) NOT(Behav* AND employee*) WN KY

NOT ((Behavior AND employees) WN ALL))

FIGURE .4 Commande de recherche d'experts utilisée dans "Engineering Village"

1. **Définition des critères de recherche :** Établissement des critères pour sélectionner les articles pertinents à examiner, tels que la pertinence du sujet, la qualité de la recherche et la date de publication.
2. **Recherche de la littérature :** Identification des sources d'information pertinentes et recherche systématique d'articles correspondant aux critères de recherche établis. Les bases de données académiques, les revues spécialisées et les conférences ont été consultées pour collecter les articles pertinents.
3. **Sélection des études :** Tri des articles trouvés en fonction de leur pertinence pour les questions de recherche et de leur qualité. Les articles sélectionnés ont été soumis à un examen complet pour extraire les informations nécessaires.
4. **Extraction et analyse des données :** Compilation des informations recueillies à partir des articles sélectionnés et analyse des métriques de conformité informatique existantes. Les métriques ont été classées en fonction de leur pertinence pour les questions de recherche et de leur applicabilité dans divers contextes.
5. **Synthèse des résultats :** Intégration des métriques identifiées dans la BDMC et évaluation de leur pertinence pour répondre aux questions de recherche. Les métriques validées ont été ajoutées à la base de données pour aider à quantifier le niveau de conformité et à identifier les points faibles.

ANNEXE 2: STRUCTURE DE L'ONTOLOGIE

La Figure .5 illustre l'organisation hiérarchique des classes et des individus de l'ontologie représenté sous l'Éditeur Protégé.

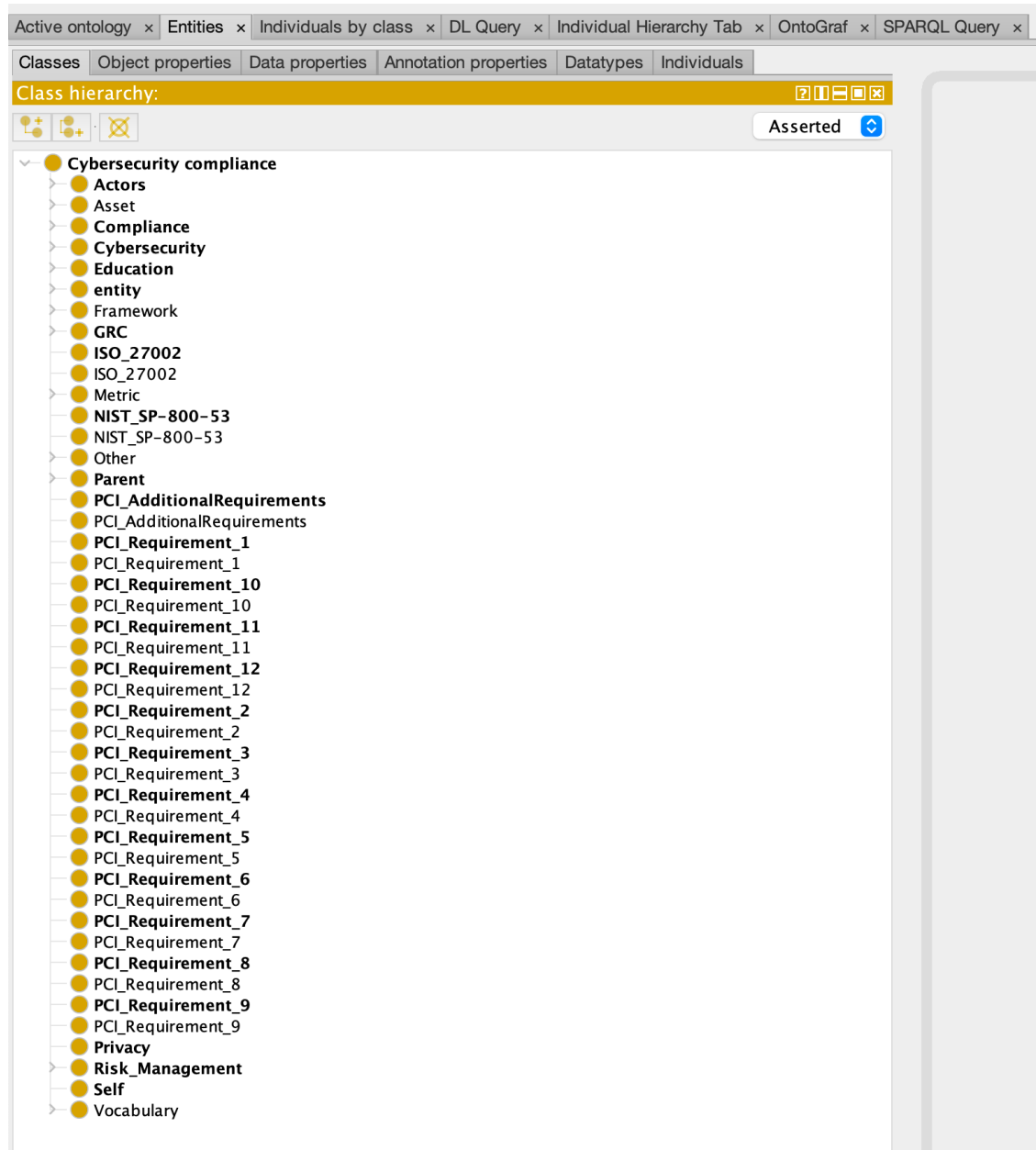


FIGURE .5 Hiérarchie des classes et individus de l'ontologie.

Par ailleurs, l'ontologie peut être divisée en quatre parties, chacune se concentrant sur différents aspects du domaine, ce qui permet une représentation plus gérable et plus compréhensible des concepts et de leurs relations. Les quatre parties sont les suivantes :

1. Les classes liées aux actifs, aux menaces, aux vulnérabilités et aux contrôles (Cf. Figure .6)
2. Classes relatives aux utilisateurs et au contrôle d'accès (Cf. Figure .7)
3. Classes relatives au contrôle de la sécurité, à l'architecture et aux données (Cf. Figure .8)
4. Classes relatives à la vie privée et à la réglementation (Cf. Figure .9)

La figure .6 illustre les classes liées aux actifs, aux menaces, aux vulnérabilités et aux contrôles. Elle met en évidence les concepts et les relations fondamentaux qui permettent de comprendre les risques de sécurité et leurs stratégies d'atténuation, jetant ainsi les bases de la gestion de la sécurité au sein d'une organisation.

La figure .7 se concentre sur les classes liées aux utilisateurs et au contrôle d'accès. Ce diagramme souligne l'importance de la confidentialité dans la gestion des utilisateurs et le contrôle de l'accès aux ressources, en veillant à ce que les informations sensibles soient protégées contre tout accès non autorisé et toute utilisation abusive. Le diagramme montre les concepts clés et les relations nécessaires pour gérer efficacement les utilisateurs, les rôles et les droits d'accès, qui sont essentiels pour maintenir la sécurité et la confidentialité des données d'une organisation.

La figure .8 présente les classes relatives au contrôle de la sécurité, à l'architecture et aux données. Cette partie de l'ontologie illustre le lien entre les contrôles de sécurité et l'architecture globale de l'organisation, ainsi que les différents types de données impliqués dans les processus de sécurité. Il est essentiel de comprendre ces relations pour mettre en œuvre des contrôles de sécurité appropriés aux niveaux organisationnel et technique.

Enfin, la figure .9 traite des classes liées à la vie privée et à la réglementation. Ce diagramme souligne l'importance du respect des diverses réglementations en matière de protection de la vie privée et présente les concepts et les relations qui aident les organisations à se conformer à ces exigences. Il souligne également le rôle des mécanismes de protection de la vie privée et leur impact sur la protection des données.

La présence de classes dupliquées dans les quatre diagrammes est intentionnelle, car elles servent à fournir un contexte et à démontrer les relations entre les différentes classes dans chaque diagramme. Il est important de noter que, dans l'ontologie complète, chaque classe n'existe qu'une seule fois, et que les diagrammes plus petits sont simplement des vues différentes du modèle plus large, mettant l'accent sur des sous-ensembles spécifiques de classes et de relations. En divisant l'ontologie en diagrammes plus petits, il devient plus facile de comprendre et d'analyser des domaines d'intérêt spécifiques sans être submergé par la complexité de l'ensemble du modèle.

Par ailleurs, la protection de la vie privée est un aspect important de la gestion des utilisateurs et du contrôle de l'accès aux ressources, car elle garantit que les informations sensibles sont protégées contre tout accès non autorisé et toute utilisation abusive. Lorsque les organisations collectent et traitent des données personnelles, elles ont la responsabilité de les protéger contre les failles de sécurité et de se conformer aux diverses réglementations en matière de protection de la vie privée. La mise en œuvre correcte des mécanismes de gestion des utilisateurs et de contrôle d'accès est essentielle pour empêcher l'accès non autorisé à des informations privées, ce qui peut entraîner des violations de données, des pertes financières et une atteinte à la réputation d'une organisation. En outre, en maintenant un niveau élevé de protection de la vie privée, les organisations peuvent établir une relation de confiance avec leurs utilisateurs, leurs clients et leurs partenaires.

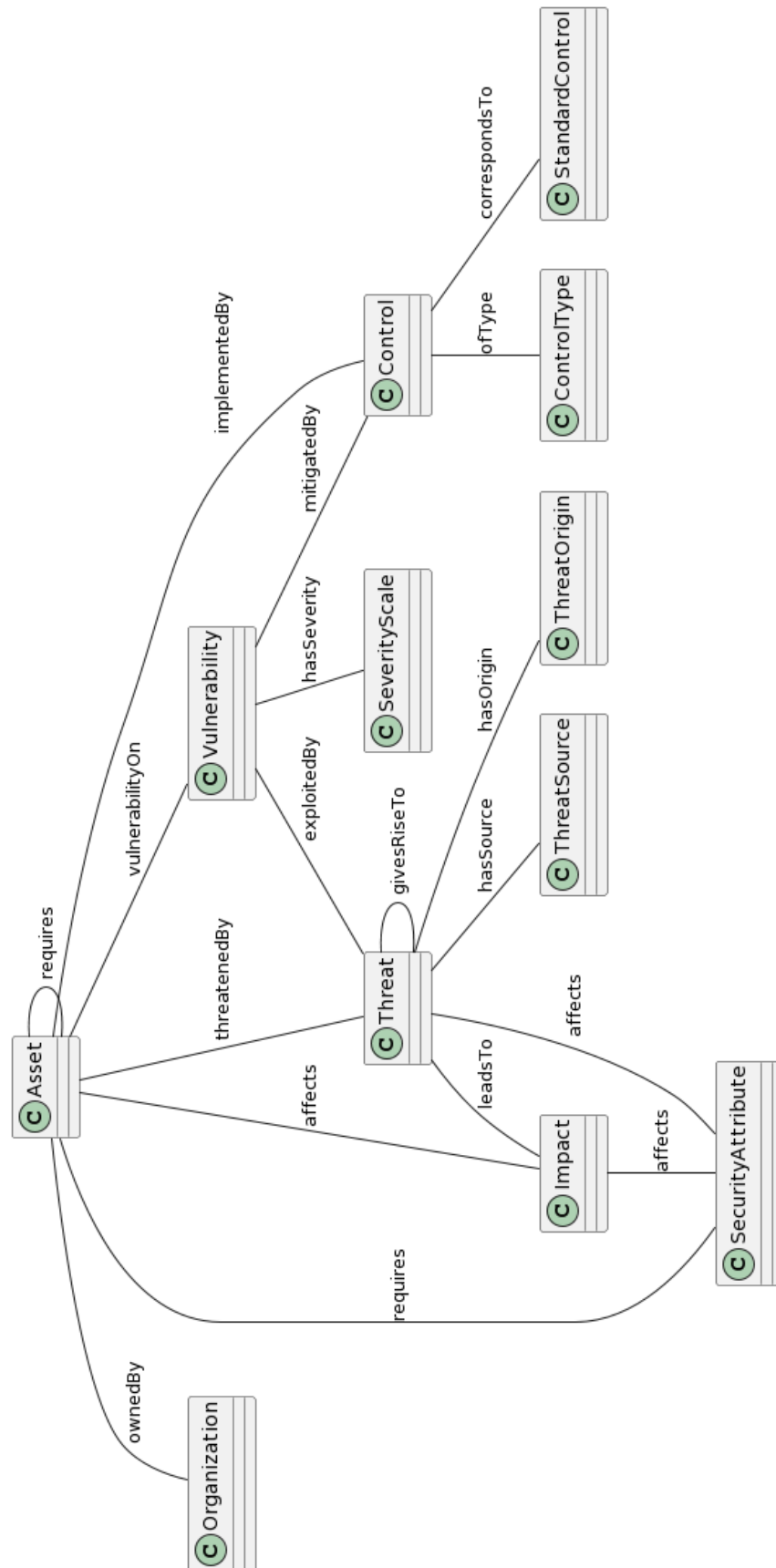


FIGURE .6 Classes liées aux actifs, aux menaces, aux vulnérabilités et aux contrôles

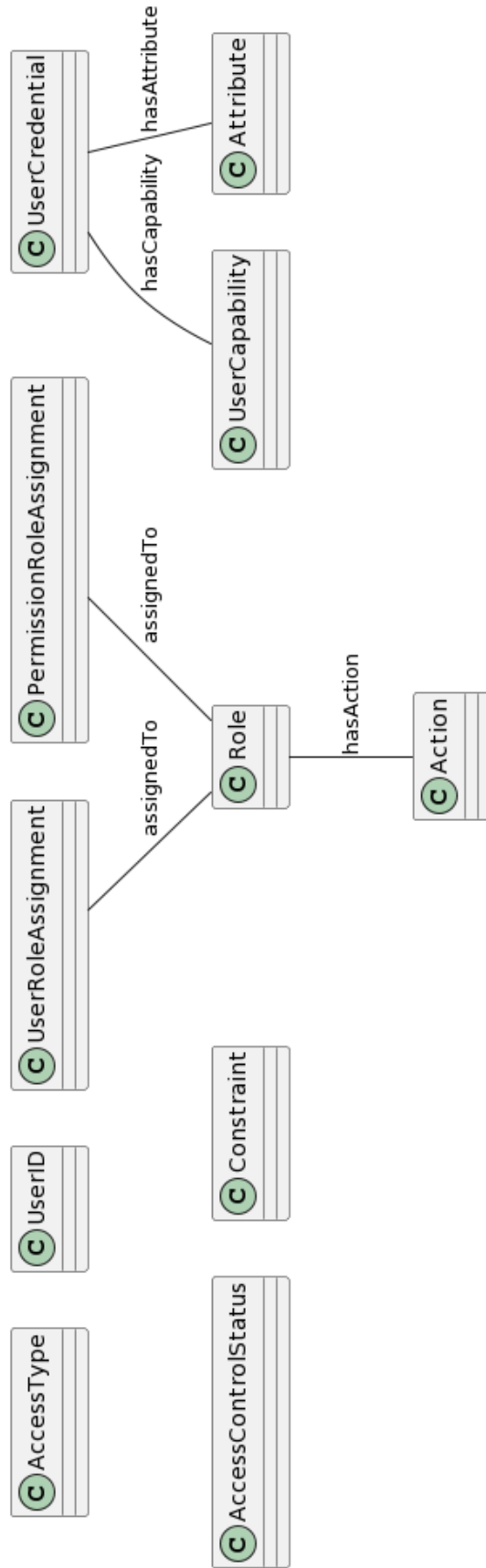


FIGURE .7 Classes liées aux utilisateurs et au contrôle d'accès

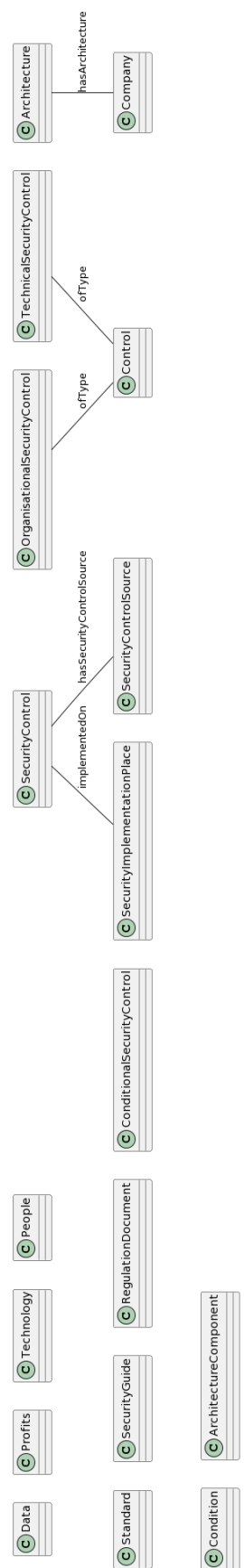


FIGURE .8 Contrôle de sécurité, architecture et classes de données

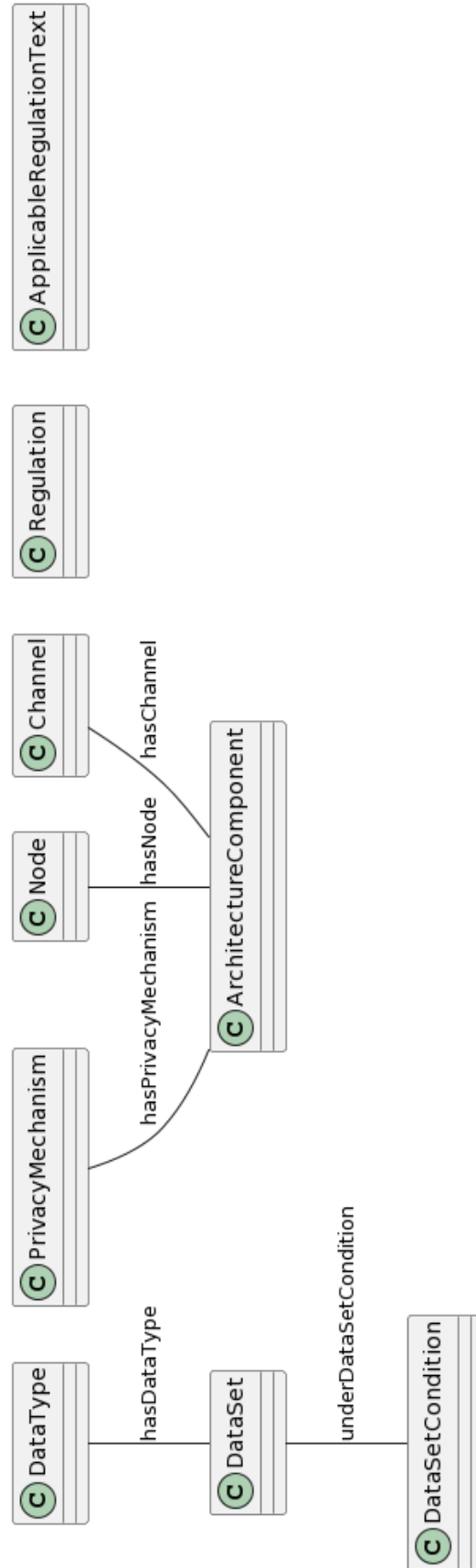


FIGURE .9 Classes liées à la protection de la vie privée et à la réglementation