UQAC
Université du Québec
à Chicoutimi

CLASSIFICATION CONTEXTUELLE DES ALERTES DANS LES CENTRES D'OPÉRATIONS DE SÉCURITÉ À L'AIDE DES GRANDS MODÈLES DE LANGAGE

PAR ELHADJ ABDOURAHMANE BALDE

MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI COMME EXIGENCE PARTIELLE EN VUE DE L'OBTENTION DU GRADE DE MAÎTRE ÈS SCIENCES EN INFORMATIQUE

QUÉBEC, CANADA

RÉSUMÉ

Les centres d'opérations de sécurité (Security Operations Centers, SOC) traitent quotidiennement un volume élevé d'alertes générées par les plateformes de gestion des informations et des événements de sécurité (Security Information and Event Management, SIEM), dont une proportion importante correspond à des faux positifs. Cette situation impose une charge significative aux analystes de niveau Tier-1 lors du processus de triage des alertes. En pratique, ces analystes s'appuient fortement sur le contexte organisationnel, notamment la criticité des actifs, les rôles des utilisateurs, l'environnement réseau, les plages horaires et les politiques de sécurité, afin d'interpréter et de prioriser correctement les alertes. Toutefois, les tentatives visant à intégrer directement ces informations contextuelles dans les règles de détection SIEM ou dans des logiques de classification en aval entraînent souvent une duplication de la logique, des mises à jour fréquentes à mesure que les environnements évoluent, et des limites en matière de scalabilité.

Dans ce contexte, l'intelligence artificielle, et plus particulièrement les modèles de langage de grande taille (Large Language Models, LLMs), apparaît comme une piste prometteuse pour assister le triage des alertes SOC. Grâce à leur capacité à analyser des descriptions textuelles, à traiter des informations hétérogènes et à produire des justifications compréhensibles, les LLMs peuvent soutenir les analystes Tier-1 dans l'évaluation initiale des alertes. Toutefois, leur utilisation dans un environnement SOC soulève plusieurs limites. En l'absence de contexte organisationnel explicite, ces modèles peuvent produire des décisions conservatrices, générer des faux positifs ou formuler des justifications plausibles mais incomplètes. De plus, leur nature probabiliste peut entraîner une variabilité des réponses lors d'exécutions répétées, ce qui pose un enjeu de fiabilité pour une utilisation opérationnelle.

Ce mémoire évalue la capacité d'un modèle de langage de grande taille (Large Language Model, LLM), Meta LLaMA 3.3–70B-Instruct-Turbo, à faire le triage des alertes dans un contexte SOC Tier-1 en intégrant explicitement le contexte organisationnel, tout en dissociant la gestion du contexte de la logique de classification encodée dans l'invite du modèle. L'objectif n'est pas de mesurer la performance générale des LLMs sur un grand volume d'alertes, mais plutôt d'isoler l'effet du contexte organisationnel sur les décisions de triage dans des conditions expérimentales contrôlées.

L'environnement expérimental repose sur la plateforme SIEM open source Wazuh, surveillant des hôtes Windows et Ubuntu soumis à des scénarios d'émulation d'attaques générés à l'aide de CALDERA, ainsi qu'à des activités système légitimes. Un corpus d'alertes annotées manuellement est utilisé pour comparer deux configurations expérimentales : une première fondée uniquement sur les métadonnées SIEM brutes, et une seconde dans laquelle le triage des alertes est enrichi par l'injection directe du contexte organisationnel dans l'invite du modèle, sous la forme d'un fichier JSON structuré.

En l'absence de contexte organisationnel, le modèle atteint une précision globale de 85,45 %, avec une précision de 80 %, un rappel de 96,55 % et un F1-score de 87,5 %. L'analyse des erreurs montre que les faux positifs observés sont principalement liés à l'incapacité du modèle à distinguer certains événements bénins d'activités réellement suspectes lorsque les informations organisationnelles sont absentes. Lorsque le contexte organisationnel est intégré, le modèle ne génère aucun faux positif sur le corpus évalué et produit des classifications stables lors des exécutions répétées réalisées dans les mêmes conditions expérimentales.

Ces résultats montrent que l'intégration explicite du contexte organisationnel peut améliorer la qualité du triage des alertes par les LLMs, en réduisant les faux positifs et en renforçant la stabilité des classifications. Le mémoire met ainsi en évidence l'importance du contexte pour une utilisation plus fiable des modèles de langage dans les opérations SOC.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

TP *True Positives* : Alertes intéressantes correctement détectées

TN *True Negatives* : Alertes bénignes correctement ignorées

FP *False Positives* : Alertes bénignes incorrectement signalées comme intéressantes

FN *False Negatives* : Alertes intéressantes non détectées par le modèle

TPR *True Positive Rate* : Taux de vrais positifs (équivalent au rappel)

FPR *False Positive Rate* : Taux de faux positifs

F1-score *F1-score* : Moyenne harmonique de la précision et du rappel

SOC *Security Operations Center* : Centre d'opérations de sécurité

SIEM *Security Information and Event Management* : Plateforme centralisée de collecte et de corrélation des événements de sécurité

IDS *Intrusion Detection System* : Système de détection d'intrusion

IPS *Intrusion Prevention System* : Système de prévention des intrusions

TTP *Tactics, Techniques and Procedures* : Tactiques, techniques et procédures d'attaque (MITRE ATT&CK)

LLM *Large Language Model* : Modèle de langage de grande taille

IAM *Identity and Access Management* : Gestion des identités et des accès

CMDB *Configuration Management Database* : Base de données de gestion de configuration

GRC *Governance, Risk and Compliance* : Gouvernance, gestion des risques et conformité

JSON *JavaScript Object Notation* : Format structuré de représentation des données

API *Application Programming Interface* : Interface de programmation applicative

VM *Virtual Machine* : Machine virtuelle

DÉDICACE

À la mémoire de mon frère, décédé le 9 avril 2025, pour son soutien constant et ses encouragements indéfectibles.

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de mémoire Jonathan Roy pour son encadrement, sa disponibilité et ses conseils tout au long de ce travail. Son expertise et ses orientations ont été essentielles à la réalisation de ce mémoire.

Je remercie également les membres de l'Université du Québec à Chicoutimi (UQAC) pour l'environnement académique et les ressources mises à disposition, qui ont permis de mener cette recherche dans de bonnes conditions.

Je souhaite aussi remercier les organisations et plateformes techniques ayant rendu possible ce travail expérimental, en particulier le Cyber-Québec cyber range, utilisé comme environnement de laboratoire et d'expérimentation en cybersécurité dans le cadre de cette recherche.

Enfin, je remercie sincèrement ma famille et mes proches pour leur soutien moral constant, leur patience et leurs encouragements tout au long de mon parcours universitaire.

AVANT-PROPOS

Ce mémoire s'inscrit dans le cadre du programme de maîtrise en informatique de l'Université du Québec à Chicoutimi. Il a été réalisé dans une perspective de recherche appliquée, avec pour objectif d'étudier l'apport des LLMs au triage des alertes dans les centres d'opérations de sécurité.

L'intérêt pour ce sujet est né de l'observation des défis croissants rencontrés par les équipes SOC lors du triage des alertes, notamment l'augmentation du volume d'alertes générées par les plateformes SIEM, la présence de nombreux faux positifs et la charge cognitive imposée aux analystes de premier niveau. Dans ce contexte, les LLMs offrent des possibilités intéressantes pour assister l'analyse initiale des alertes, mais leur utilisation soulève également des enjeux liés à la fiabilité des décisions, à leur stabilité et à l'intégration du contexte organisationnel.

Ce mémoire adopte une approche empirique permettant d'analyser la précision, la réduction des faux positifs et la répétabilité des classifications produites par le modèle, contribuant à une meilleure compréhension du rôle du contexte dans l'utilisation des LLMs en cybersécurité.

Ce travail a donné lieu à un article scientifique publié dans les actes de la conférence 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), tenue à Houston, Texas, du 18 au 20 février 2026 Roy & Balde (2026). Cette publication témoigne de l'intérêt scientifique du sujet, notamment pour l'utilisation des modèles de langage de grande taille dans le triage contextuel des alertes SOC.

INTRODUCTION

La cybersécurité occupe aujourd'hui une place centrale dans le fonctionnement des organisations modernes. La dépendance croissante aux systèmes numériques, aux services infonuagiques et aux infrastructures interconnectées augmente la quantité de journaux, d'événements et d'alertes de sécurité à analyser. Dans ce contexte, la capacité à détecter, analyser et prioriser rapidement les alertes devient un enjeu stratégique pour limiter les risques d'incidents et améliorer la réponse opérationnelle.

Pour détecter les activités suspectes, les tentatives d'intrusion et les incidents de sécurité potentiels, les organisations s'appuient sur des centres d'opérations de sécurité (Security Operations Centers, SOCs). Ces centres sont chargés de surveiller en continu les environnements informatiques et de traiter les alertes générées par différents outils de sécurité. Toutefois, l'augmentation constante du volume d'alertes complique cette mission. Les analystes sont confrontés à une surcharge d'alertes, dont une grande partie ne correspond pas à des incidents réels, ce qui engendre une pression opérationnelle importante et affecte la qualité du triage.

Dans ce contexte, l'automatisation partielle du traitement des alertes apparaît comme une piste prometteuse pour soutenir les équipes SOC. Les avancées récentes en intelligence artificielle, et plus particulièrement les modèles de langage de grande taille (Large Language Models, LLMs), ouvrent de nouvelles perspectives pour l'assistance au triage des alertes. Ces modèles offrent la possibilité d'interpréter des alertes, de produire des justifications compréhensibles et d'appuyer le travail des analystes.

Cependant, l'intégration de ces modèles dans les processus SOC soulève plusieurs questions. Le triage effectué par un analyste humain ne repose pas uniquement sur des indicateurs techniques, mais également sur une compréhension fine du contexte organisationnel, incluant les rôles des utilisateurs, les plages horaires normales d'activité, la criticité des actifs, les

contraintes métiers et les politiques de sécurité. En l'absence de ces informations, un LLM peut interpréter une alerte bénigne comme suspecte ou, inversement, sous-estimer une alerte nécessitant une investigation.

Ce mémoire s'inscrit dans cette problématique et vise à étudier l'apport du contexte organisationnel dans l'utilisation des modèles de langage pour le triage des alertes de sécurité. L'objectif général est d'analyser dans quelle mesure l'intégration explicite de ce contexte peut améliorer la classification des alertes, réduire les faux positifs et renforcer la cohérence des réponses produites par un LLM dans un environnement SOC contrôlé.

CHAPITRE I

INTRODUCTION

Les centres d'opérations de sécurité (Security Operations Centers, SOCs) constituent aujourd'hui un pilier central des stratégies de cybersécurité des organisations. Leur rôle est d'assurer une surveillance continue des systèmes d'information, de détecter les activités suspectes et de coordonner la réponse aux incidents de sécurité. Cette mission repose principalement sur des outils tels que les systèmes de détection et de prévention d'intrusion (Intrusion Detection Systems – IDS, Intrusion Prevention Systems – IPS) ainsi que les plateformes de gestion des événements et informations de sécurité (Security Information and Event Management, SIEM).

Toutefois, l'augmentation constante du volume d'événements collectés par ces systèmes a conduit à une multiplication des alertes de sécurité, dont une proportion importante correspond à des faux positifs. Ce phénomène engendre une fatigue liée aux alertes (alert fatigue), qui affecte directement l'efficacité opérationnelle des analystes SOC et la qualité de leurs décisions, en particulier lors du triage des alertes Alahmadi *et al.* (2022). Plusieurs travaux soulignent que cette surcharge informationnelle constitue l'un des défis majeurs des SOC modernes, notamment en raison du volume élevé d'alertes, du manque de temps disponible pour les analyser et de la proportion importante de faux positifs Zhong *et al.* (2016); Radu *et al.* (2025).

Au-delà du volume d'alertes, la littérature met également en évidence une limite importante des processus actuels de triage : l'absence de contexte organisationnel explicite lors de la classification des alertes Alahmadi *et al.* (2022); Tariq *et al.* (2025). Les alertes générées par les plateformes SIEM reposent principalement sur des indicateurs techniques de

bas niveau et omettent fréquemment des éléments contextuels essentiels à une interprétation correcte. Ces éléments incluent notamment les calendriers opérationnels (processus de fin de mois, sauvegardes, audits), les schémas d'activité des utilisateurs (télétravail, déplacements, tâches planifiées), ainsi que la distinction entre périodes normales et anormales d'activité. En pratique, les analystes Tier-1 s'appuient sur ces connaissances contextuelles afin d'interpréter les alertes au-delà de leurs seuls indicateurs techniques et de distinguer les faux positifs, les événements bénins et les alertes réellement pertinentes nécessitant une investigation Alahmadi *et al.* (2022).

Lorsque ces informations contextuelles sont absentes, les analystes compensent par une classification manuelle fondée sur leur expérience et sur des connaissances organisationnelles implicites. Les tentatives visant à intégrer ce contexte directement dans les règles de détection des SIEM se heurtent toutefois à des limites importantes, telles que la duplication de la logique de détection, la nécessité de mises à jour fréquentes à mesure que les environnements organisationnels évoluent, et une faible capacité de passage à l'échelle. Zhong et al. Zhong *et al.* (2016) ainsi que Radu et al. Radu *et al.* (2025) montrent que les analystes Tier-1 opèrent sous une forte contrainte temporelle et avec une information incomplète : seule une fraction des alertes générées est effectivement analysée, et une part significative de celles-ci correspond à des faux positifs, illustrant l'ampleur du problème opérationnel.

Dans ce contexte, l'automatisation partielle du triage des alertes apparaît comme une piste prometteuse pour soutenir les équipes SOC. Les avancées récentes en intelligence artificielle, et en particulier en modèles de langage de grande taille (Large Language Models, LLMs), ouvrent de nouvelles perspectives pour l'assistance à la décision dans des tâches complexes impliquant l'analyse de données textuelles. Des solutions industrielles, telles que les approches d'assistance basées sur l'IA proposées pour les SOC Freitas *et al.* (2025), témoignent de l'intérêt croissant pour ces technologies. Les LLMs sont particulièrement

adaptés au triage des alertes en raison de leur capacité à interpréter des descriptions textuelles non structurées et à produire des justifications compréhensibles, proches du raisonnement humain Bono *et al.* (2024).

Néanmoins, l'efficacité de la classification des alertes basée sur les LLMs dépend fortement de l'ancrage contextuel. Lorsqu'ils opèrent uniquement à partir des métadonnées SIEM brutes, les LLMs ont tendance à mal classifier les alertes, en particulier dans des scénarios ambigus où les indicateurs techniques seuls sont insuffisants pour déterminer l'intention réelle Tariq *et al.* (2025). Oniagbi Oniagbi (2024) a montré une amélioration de la précision de classification lorsque des informations contextuelles étaient fournies, mais cette évaluation restait limitée à une seule instance d'alerte et n'examinait ni la stabilité ni la répétabilité des décisions du modèle.

Ces constats mettent en évidence une lacune importante dans les travaux existants. Bien que l'importance du contexte organisationnel dans les opérations SOC soit largement reconnue, son rôle dans la classification des alertes basée sur les LLMs n'a pas été systématiquement modélisé ni évalué comme variable expérimentale explicite. En particulier, l'impact du contexte sur la précision de classification, la réduction des faux positifs et la répétabilité des décisions demeure insuffisamment compris.

Dans ce cadre, ce mémoire vise à évaluer empiriquement l'influence du contexte organisationnel sur la classification des alertes basée sur les LLMs dans des conditions contrôlées. L'objectif n'est pas d'estimer des performances généralisables à grande échelle, mais d'isoler l'effet spécifique du contexte sur les résultats du triage des alertes. Pour ce faire, les décisions de classification obtenues avec et sans contexte organisationnel sont comparées afin d'analyser leur impact sur la précision de classification, le taux de faux positifs et la répétabilité des décisions sur plusieurs exécutions dans des conditions identiques.

Pour soutenir cette investigation, un environnement SOC expérimental a été mis en place à l'aide de la plateforme SIEM open source Wazuh et du cadre d'émulation d'adversaires CALDERA, surveillant des systèmes Windows et Linux soumis à des activités adverses et légitimes. Les alertes produites ont été annotées manuellement afin d'établir une vérité terrain, puis analysées à l'aide du modèle Meta LLaMA 3.3–70B-Instruct-Turbo. Les résultats montrent une amélioration significative de la qualité du triage, une réduction des faux positifs et une meilleure cohérence des décisions lorsque le contexte organisationnel est pris en compte.

Enfin, bien que cette étude repose sur une représentation statique du contexte sous forme de fichier structuré, elle ouvre la voie à des travaux futurs visant à intégrer des sources de contexte dynamiques, telles que les bases de gestion de configuration (CMDB), les systèmes de gestion des identités et des accès (IAM) ou encore les outils de gouvernance, de gestion des risques et de conformité (GRC).

1.0.1 PROBLÉMATIQUE

La multiplication des alertes de sécurité au sein des centres d'opérations de sécurité (Security Operations Centers, SOCs) constitue aujourd'hui un défi opérationnel majeur. Cette situation résulte à la fois de l'augmentation de la surface d'attaque des organisations et de la généralisation des plateformes de détection telles que les IDS, IPS et SIEM, qui génèrent un volume élevé d'alertes hétérogènes. Une part importante de ces alertes correspond à des faux positifs ou à des événements bénins, contribuant à une fatigue liée aux alertes susceptible de dégrader la qualité du triage et d'augmenter le risque d'erreurs humaines Alahmadi *et al.* (2022).

Cette problématique s'inscrit dans un contexte plus large marqué par l'essor récent de l'intelligence artificielle, et en particulier des modèles de langage de grande taille (Large

Language Models, LLMs). En cybersécurité, plusieurs travaux ont mis en évidence le potentiel de l'IA pour améliorer la détection d'incidents, l'analyse des menaces et l'assistance à la prise de décision Zeadally *et al.* (2020). Toutefois, malgré ces avancées scientifiques, l'adoption des solutions basées sur l'IA dans les environnements opérationnels demeure relativement limitée. Motlagh et al. soulignent que la cybersécurité est historiquement lente à adopter des approches fondées sur les données et l'apprentissage automatique Motlagh *et al.* (2024), un constat également confirmé par des rapports industriels récents indiquant qu'une minorité d'organisations exploitent effectivement des outils de sécurité reposant sur l'IA Gartner, Inc. (2023).

Cette adoption limitée intervient alors même que les SOCs font face à des contraintes structurelles croissantes, notamment une pénurie mondiale de compétences en cybersécurité. Plusieurs millions de postes restent vacants à l'échelle internationale, y compris pour des rôles d'entrée tels que les analystes SOC de niveau Tier-1 Furnell (2021). Cette pénurie accentue la pression exercée sur les équipes en place, qui doivent analyser un volume croissant d'événements de sécurité avec des ressources humaines limitées, contribuant ainsi à la surcharge cognitive et au risque d'épuisement professionnel Sundaramurthy *et al.* (2015).

Au-delà du volume d'alertes, la littérature met en évidence une limite structurelle des systèmes actuels de détection : l'absence de contexte organisationnel explicite lors de la classification des alertes Alahmadi *et al.* (2022); Tariq *et al.* (2025). Les alertes SIEM reposent majoritairement sur des indicateurs techniques de bas niveau et omettent des éléments essentiels tels que les calendriers opérationnels, les habitudes d'activité des utilisateurs, les contraintes métiers, la criticité des actifs ou encore la distinction entre périodes normales et anormales d'activité. En pratique, les analystes SOC de niveau Tier-1 mobilisent ces informations contextuelles de manière implicite pour interpréter les alertes et orienter leurs décisions.

Bien que les plateformes SIEM permettent l'intégration de certaines formes de contexte organisationnel, cette intégration repose majoritairement sur des mécanismes statiques et explicites, qui deviennent difficiles à maintenir et à faire évoluer dans des environnements opérationnels complexes et dynamiques.

La Figure 1.1 illustre une règle SIEM dans laquelle le contexte organisationnel est directement intégré à la logique de détection. Cette règle vise à détecter des accès à des ressources sensibles en dehors des heures ouvrables par des utilisateurs non autorisés. Dans ce cas, des informations contextuelles telles que la liste des ressources sensibles, les utilisateurs autorisés et les plages horaires normales doivent être encodées explicitement dans la règle.

```
let SensitiveResources = dynamic(["FIN-DB01", "HR-DB02"]);
let AuthorizedUsers = dynamic(["dba_john", "dba_sara", "dba_service"]);
let BusinessHours = 08 .. 18;

AuditLogs
| where Action == "Access"
| where TargetResource in (SensitiveResources)
| where UserName !in (AuthorizedUsers)
| where datetime_part("hour", Timestamp) !in (BusinessHours)
| summarize count() by UserName, TargetResource, Timestamp
```

FIGURE 1.1 : Exemple de règle SIEM intégrant directement le contexte organisationnel

Comme le montre la Figure 1.1, le contexte organisationnel est directement intégré à la logique de détection. Ainsi, toute modification de l'environnement, comme l'ajout d'une nouvelle ressource sensible, l'arrivée d'un nouvel utilisateur autorisé ou un changement des heures ouvrables, impose une mise à jour manuelle de la règle. À grande échelle, cette approche entraîne une duplication de la logique contextuelle dans plusieurs règles et complique la maintenance du SIEM.

Cette difficulté technique a des conséquences directes sur le triage opérationnel des alertes. Lorsque les règles ne reflètent pas correctement l'évolution du contexte organisationnel, elles peuvent générer davantage d'alertes ambiguës ou faussement positives que les analystes doivent ensuite interpréter manuellement.

Zhong *et al.* (2016) ainsi que Radu *et al.* (2025) montrent que les analystes Tier-1 opèrent sous une forte contrainte temporelle et avec une information incomplète : seule une fraction des alertes générées est effectivement analysée, et une proportion significative de celles-ci correspond à des faux positifs. Ces constats illustrent la difficulté de maintenir un triage fiable dans des conditions opérationnelles réalistes.

Dans ce contexte, les modèles de langage de grande taille ont suscité un intérêt croissant pour l'automatisation ou l'assistance au triage des alertes SOC. Leur capacité à analyser des descriptions textuelles complexes et à produire des justifications compréhensibles en fait des outils particulièrement prometteurs pour assister les analystes SOC de niveau 1 (Tier-1) Bono *et al.* (2024). Plusieurs travaux récents indiquent que l'intégration d'informations contextuelles peut améliorer la qualité des classifications produites par ces modèles Tariq *et al.* (2025); Oniagbi (2024).

Cependant, les LLMs présentent des limites intrinsèques liées à leur nature probabiliste. À paramètres et entrées identiques, un même modèle peut produire des sorties différentes d'une exécution à l'autre, un phénomène souvent associé aux hallucinations ou à l'instabilité des décisions. Dans un environnement critique tel qu'un SOC, cette variabilité soulève des enjeux majeurs de fiabilité et de confiance, puisqu'une même alerte pourrait être classifiée différemment selon l'exécution du modèle. Or, la majorité des travaux existants se concentrent principalement sur la précision de classification ou la réduction des faux positifs, sans examiner systématiquement la stabilité des décisions produites par les LLMs.

La notion de répétabilité, définie comme la capacité d'un modèle à produire des sorties de signification cohérente lorsqu'un même cas est évalué à plusieurs reprises dans des conditions strictement identiques, constitue pourtant un critère essentiel pour une intégration opérationnelle des LLMs Shyr *et al.* (2025). Sans une stabilité minimale des décisions, l'automatisation du triage des alertes risque d'introduire une nouvelle source d'incertitude plutôt que de réduire la charge cognitive des analystes.

Ainsi, malgré les avancées récentes dans l'application des LLMs au triage des alertes SOC, plusieurs lacunes subsistent : (i) le rôle du contexte organisationnel est rarement modélisé comme une variable expérimentale explicite, (ii) l'impact de ce contexte sur la réduction des faux positifs demeure insuffisamment quantifié, et (iii) la répétabilité des décisions produites par les LLMs reste largement sous-explorée. Ces limites motivent la nécessité d'une évaluation expérimentale contrôlée visant à analyser conjointement l'effet du contexte organisationnel et la stabilité des décisions dans un cadre SOC réaliste.

1.0.2 OBJECTIFS

L'objectif principal de ce mémoire est d'analyser le rôle du contexte organisationnel dans le triage des alertes au sein des centres d'opérations de sécurité (SOC) à l'aide des modèles de langage de grande taille (Large Language Models, LLMs). Contrairement aux travaux visant à comparer les performances de plusieurs modèles, cette étude cherche à isoler et à caractériser l'influence spécifique du contexte organisationnel sur le comportement décisionnel d'un LLM, dans un cadre expérimental contrôlé.

Plus précisément, ce mémoire vise à :

— Étudier la capacité d'un LLM à réaliser le triage des alertes SOC au niveau Tier-1 à partir des métadonnées produites par une plateforme SIEM.

— Analyser les différences de décisions produites par le modèle en présence et en absence de contexte organisationnel explicite.
— Évaluer l'impact du contexte organisationnel sur la précision de classification et sur la réduction des faux positifs.
— Examiner la stabilité et la répétabilité des décisions du modèle lorsque les mêmes alertes sont évaluées à plusieurs reprises dans des conditions strictement identiques.

1.0.3 QUESTIONS DE RECHERCHE

Afin de répondre à ces objectifs, ce mémoire s'articule autour des questions de recherche suivantes :

— **RQ1** : Comment un modèle de langage de grande taille peut-il être utilisé pour réaliser le triage des alertes SOC au niveau Tier-1 à partir de données SIEM ?
— **RQ2** : Quel est l'impact de l'intégration d'un contexte organisationnel explicite sur la précision de classification et le taux de faux positifs des décisions produites par un LLM ?
— **RQ3** : Dans quelle mesure le contexte organisationnel influence-t-il la stabilité et la répétabilité des décisions de triage produites par un LLM lors d'exécutions répétées dans des conditions identiques ?

1.0.4 STRUCTURE DU MÉMOIRE

Le reste de ce mémoire est organisé comme suit. Le chapitre 2 présente les concepts fondamentaux liés aux centres d'opérations de sécurité, aux processus de détection et de triage des alertes, ainsi qu'aux plateformes SIEM. Le chapitre 3 introduit les modèles de langage de grande taille, leur fonctionnement général et leurs applications en cybersécurité,

en mettant l'accent sur leurs limites dans les contextes opérationnels SOC. Le chapitre 4 décrit l'environnement expérimental mis en place, la méthodologie adoptée et l'architecture du système utilisé pour évaluer l'impact du contexte organisationnel sur le triage des alertes. Le chapitre 5 présente et analyse les résultats expérimentaux, tant sur le plan quantitatif que qualitatif, en comparant les performances du modèle avec et sans contexte organisationnel, ainsi que la répétabilité des décisions produites. Enfin, le chapitre 6 conclut le mémoire en synthétisant les principales contributions de cette étude, en discutant ses limites et en proposant des perspectives de recherche futures.

CHAPITRE II

CENTRES D'OPÉRATIONS DE SÉCURITÉ ET TRIAGE DES ALERTES

Ce chapitre présente les fondements organisationnels et opérationnels des centres d'opérations de sécurité. Il décrit le rôle des SOC, les plateformes SIEM, le processus de triage des alertes au niveau Tier-1 ainsi que les limites associées à la surcharge d'alertes et au triage humain.

2.1 DÉFINITION

Un centre d'opérations de sécurité (Security Operations Center, SOC) constitue une entité centralisée chargée d'assurer la surveillance continue des systèmes d'information d'une organisation afin de détecter, analyser, prioriser et répondre aux incidents de sécurité. Le SOC joue un rôle central dans la posture de cybersécurité organisationnelle en servant de point de convergence pour les événements de sécurité issus de multiples sources hétérogènes Mutemwa *et al.* (2018); Vielberth & Pernul (2018).

Les services SOC peuvent être assurés selon différents modèles organisationnels. Certaines organisations disposent d'un SOC interne entièrement dédié, tandis que d'autres externalisent tout ou partie de ces fonctions à des fournisseurs de services de sécurité managés (Managed Security Service Providers, MSSP). Des modèles hybrides, combinant une équipe interne et des services externalisés, sont également couramment observés Vielberth & Pernul (2018). Quel que soit le mode de déploiement, le SOC est généralement fortement intégré à l'infrastructure informatique et réseau de l'organisation, et agit comme un point central de collecte et d'analyse des événements de sécurité.

L'une des caractéristiques fondamentales d'un SOC est la surveillance continue des environnements numériques. Cette surveillance vise à détecter les activités suspectes en quasi temps réel afin de limiter l'impact potentiel des attaques et de permettre une réponse rapide Mughal (2022). À ce titre, plusieurs travaux soulignent que l'existence et la maturité d'un SOC constituent des indicateurs clés du niveau global de sécurité d'une organisation Mutemwa *et al.* (2018).

2.2 ORGANISATION ET RÔLES AU SEIN D'UN SOC

Les activités d'un SOC reposent sur une organisation structurée impliquant différents rôles et niveaux de responsabilité. Bien que les modèles organisationnels varient selon la taille et la maturité des organisations, une structure hiérarchique en niveaux (tiers) est largement adoptée dans la pratique Vielberth & Pernul (2018); Zhong *et al.* (2016).

Les analystes SOC sont généralement répartis en trois niveaux. Les analystes de niveau Tier-1 sont responsables du triage initial des alertes générées par les outils de sécurité. Leur mission consiste à analyser les alertes, à déterminer leur pertinence et à identifier celles susceptibles de correspondre à de véritables incidents de sécurité. Les alertes jugées pertinentes sont ensuite escaladées vers les analystes Tier-2, chargés d'approfondir l'investigation et de confirmer l'incident. Les analystes Tier-3, quant à eux, interviennent sur des tâches avancées telles que la chasse aux menaces, l'analyse forensique et la réponse aux incidents complexes.

Au-delà de cette organisation par niveaux, un SOC peut également inclure des rôles spécialisés, tels que des analystes en renseignement sur les menaces, des spécialistes en gestion des vulnérabilités, des analystes forensiques ou encore des responsables de la conformité et de la gouvernance Mughal (2022). Cette diversité de profils reflète la complexité croissante

des environnements de sécurité et la nécessité de compétences complémentaires au sein des équipes SOC.

2.3 PLATEFORMES SIEM ET GÉNÉRATION DES ALERTES

Les plateformes de gestion des événements et des informations de sécurité (Security Information and Event Management, SIEM) constituent le cœur technologique de la majorité des SOC modernes. Elles assurent la collecte, la normalisation, la corrélation et l'analyse des journaux provenant de sources multiples, telles que les systèmes d'exploitation, les applications, les équipements réseau, les solutions IDS/IPS et les outils de sécurité périmétrique Bhatt *et al.* (2014); Vielberth & Pernul (2018).

La génération d'alertes repose principalement sur des mécanismes de détection fondés sur des règles, des signatures ou des corrélations d'événements. Ces règles sont conçues pour identifier des motifs connus associés à des comportements malveillants ou anormaux.Certaines plateformes SIEM permettent également d'enrichir les alertes à l'aide d'informations contextuelles, telles que des listes d'actifs critiques, des données de renseignement sur les menaces ou encore des profils utilisateurs.

Toutefois, l'intégration du contexte organisationnel au sein des mécanismes de détection repose majoritairement sur des règles explicites et statiques. Cette approche entraîne plusieurs limitations, notamment une duplication de la logique de détection, une maintenance coûteuse des règles et une faible capacité de passage à l'échelle dans des environnements organisationnels dynamiques Zhong *et al.* (2016). Ces contraintes contribuent à un volume élevé d'alertes peu pertinentes et à un taux important de faux positifs, augmentant ainsi la charge de travail des analystes SOC.

2.4 LE TRIAGE DES ALERTES AU NIVEAU TIER-1

Le triage des alertes constitue l'une des tâches les plus critiques et les plus chronophages au sein d'un SOC. Cette responsabilité incombe principalement aux analystes de niveau Tier-1, qui doivent examiner chaque alerte générée par les plateformes SIEM afin de déterminer si elle correspond à un événement bénin ou à un incident de sécurité potentiel Kersten *et al.* (2023).

Le processus de triage repose sur l'analyse des caractéristiques de l'alerte, de son historique, des événements connexes ainsi que du contexte organisationnel implicite. Les analystes évaluent notamment la pertinence de la règle déclenchée, la criticité de l'actif concerné, le comportement habituel des utilisateurs et les activités planifiées susceptibles d'expliquer l'alerte. Cette activité requiert un raisonnement contextuel complexe, souvent basé sur l'expérience et la connaissance tacite de l'environnement surveillé.

Dans des conditions opérationnelles réalistes, les analystes Tier-1 opèrent sous une forte contrainte temporelle et avec une information partiellement incomplète. Plusieurs études montrent que seule une fraction des alertes générées est effectivement analysée et qu'une proportion significative de celles-ci correspond à des faux positifs Zhong *et al.* (2016); Radu *et al.* (2025). Ces contraintes augmentent le risque d'erreurs humaines et rendent le triage difficilement soutenable à long terme.

2.5 FATIGUE LIÉE AUX ALERTES ET LIMITES DU TRIAGE HUMAIN

La surcharge informationnelle générée par les plateformes SIEM expose les analystes SOC à un phénomène bien documenté de fatigue liée aux alertes (alert fatigue) Alahmadi *et al.* (2022). Confrontés en continu à un volume important d'alertes, souvent de faible qualité ou redondantes, les analystes peuvent progressivement devenir désensibilisés, ce qui affecte leur

vigilance, leur cohérence décisionnelle et leur capacité à identifier les incidents réellement critiques.

Plusieurs travaux ont montré que cette fatigue contribue à des taux élevés de faux positifs persistants, à une variabilité accrue des décisions de triage et à un risque de non-détection d'événements importants Zhong *et al.* (2016); Sundaramurthy *et al.* (2015). La nature répétitive et exigeante des tâches de triage accentue également la charge cognitive, favorisant l'épuisement professionnel et un taux de rotation élevé du personnel, en particulier parmi les analystes de niveau Tier-1.

Dans un contexte marqué par une pénurie mondiale de compétences en cybersécurité, ces facteurs fragilisent davantage les opérations SOC. Ils soulignent la nécessité de mécanismes d'assistance capables de réduire la charge cognitive des analystes tout en préservant, voire en améliorant, la qualité des décisions de triage.

2.6 SYNTHÈSE

Ce chapitre a présenté les fondements organisationnels et opérationnels des centres d'opérations de sécurité, ainsi que les limites inhérentes aux processus actuels de triage des alertes. Il met en évidence le rôle central du raisonnement contextuel dans les décisions des analystes SOC et les difficultés à intégrer ce raisonnement de manière évolutive au sein des plateformes SIEM.

Ces constats ouvrent la voie à l'exploration de nouvelles approches d'assistance au triage des alertes. Le chapitre suivant introduit les modèles de langage de grande taille et examine leur potentiel, ainsi que leurs limites, pour soutenir le raisonnement des analystes SOC dans un cadre contextuel.

CHAPITRE III
ÉTAT DE L'ART

Ce chapitre présente les principes généraux des modèles de langage de grande taille et leur utilisation potentielle dans les opérations SOC. Il met l'accent sur leurs applications au triage des alertes SOC, leurs limites, ainsi que sur l'importance du contexte, de l'explicabilité et de la répétabilité des classifications.

3.1 INTRODUCTION AUX MODÈLES DE LANGAGE DE GRANDE TAILLE (LLMS)

Les modèles de langage de grande taille (Large Language Models, LLMs) constituent une avancée majeure dans le domaine de l'intelligence artificielle appliquée au traitement automatique du langage naturel. Plus largement, l'intelligence artificielle vise à concevoir des systèmes capables d'analyser leur environnement et de soutenir la prise de décision humaine dans des contextes complexes Zeadally *et al.* (2020). Contrairement aux approches traditionnelles fondées sur des règles ou sur des caractéristiques définies manuellement, les LLMs reposent sur des architectures neuronales capables d'apprendre des représentations statistiques du langage à partir de volumes massifs de données textuelles.

Les avancées récentes dans ce domaine sont principalement attribuables à l'introduction des architectures de type Transformer, qui permettent de modéliser efficacement les dépendances contextuelles à longue portée au sein des séquences textuelles. Des modèles tels que GPT, PaLM ou LLaMA ont démontré des performances remarquables sur un large éventail de tâches linguistiques, ouvrant la voie à leur utilisation dans de nombreux domaines applicatifs, y compris la cybersécurité Bono *et al.* (2024).

3.2 FONCTIONNEMENT GÉNÉRAL DES LLMS

Les LLMs sont généralement entraînés selon un paradigme d'apprentissage auto-supervisé, dans lequel le modèle apprend à prédire des tokens manquants ou futurs à partir du contexte fourni. Ce processus permet au modèle d'acquérir des connaissances syntaxiques, sémantiques et, dans une certaine mesure, pragmatiques du langage.

Une fois entraînés, les LLMs peuvent être adaptés à des tâches spécifiques par le biais de techniques telles que le prompting, le few-shot learning ou le fine-tuning. Dans le cadre du triage des alertes SOC, les LLMs sont généralement sollicités à partir de descriptions textuelles d'alertes, combinées à des instructions explicites définissant le rôle attendu du modèle, par exemple celui d'un analyste SOC de niveau Tier-1. La réponse produite par le modèle correspond alors à une classification de l'alerte, souvent accompagnée d'une justification textuelle.

Dans les environnements opérationnels tels que les SOCs, cette adaptation repose principalement sur la conception d'instructions textuelles structurées, appelées prompts. La capacité à produire des explications lisibles par l'humain distingue les LLMs des approches d'apprentissage automatique traditionnelles, souvent perçues comme des boîtes noires Bono *et al.* (2024).

3.3 APPLICATIONS DES LLMS EN CYBERSÉCURITÉ

L'utilisation des LLMs en cybersécurité constitue un domaine de recherche récent mais en rapide expansion. Plusieurs travaux ont mis en évidence leur potentiel pour l'analyse de journaux de sécurité, la détection d'anomalies, l'assistance à l'investigation d'incidents et la génération de rapports Tariq *et al.* (2025); Oniagbi (2024). Ces capacités sont particulièrement

pertinentes dans les centres d'opérations de sécurité, où les analystes doivent traiter des volumes importants d'informations hétérogènes sous de fortes contraintes temporelles.

Dans le contexte des opérations SOC, les LLMs sont principalement envisagés comme des outils d'assistance à la décision destinés à soutenir les analystes humains, plutôt que comme des systèmes entièrement autonomes. Leur capacité à interpréter des descriptions textuelles complexes, à synthétiser des informations provenant de sources multiples et à produire des justifications explicites les rend particulièrement adaptés au triage des alertes au niveau Tier-1. Des solutions industrielles comme celle de Microsoft Freitas *et al.* (2025) commencent d'ailleurs à intégrer des fonctionnalités basées sur les LLMs afin d'accélérer les investigations et de réduire la charge cognitive des analystes.

Toutefois, ces applications reposent souvent sur des hypothèses implicites concernant la disponibilité et la qualité de l'information fournie au modèle, notamment en ce qui concerne le contexte organisationnel Alahmadi *et al.* (2022).

3.4 LIMITES DES LLMS APPLIQUÉS AU TRIAGE DES ALERTES SOC

Malgré leurs performances prometteuses, les LLMs présentent des limites importantes lorsqu'ils sont appliqués au triage des alertes SOC. Plusieurs études ont montré que, lorsqu'ils opèrent uniquement à partir des métadonnées SIEM brutes, les LLMs ont tendance à adopter des comportements décisionnels conservateurs Tariq *et al.* (2025); Oniagbi (2024). Dans des scénarios ambigus, le modèle privilégie fréquemment une classification indiquant un incident potentiel, contribuant ainsi à la persistance de faux positifs.

Cette limitation s'explique en grande partie par l'absence de contexte organisationnel explicite dans les données d'entrée. Contrairement aux analystes humains, les LLMs ne disposent pas d'une compréhension implicite de l'environnement opérationnel, telle que les

rôles des utilisateurs, les comportements attendus des systèmes ou les contraintes temporelles. En l'absence de ces informations, le modèle est contraint de fonder ses décisions uniquement sur des indices techniques partiels, ce qui limite la fiabilité et la stabilité de ses classifications Tariq *et al.* (2025); Oniagbi (2024).

Des travaux récents, notamment ceux d'Oniagbi Oniagbi (2024), ont montré que l'ajout de contexte organisationnel pouvait améliorer la précision des décisions produites par les LLMs. Toutefois, ces études restent limitées en termes de portée expérimentale et n'analysent pas systématiquement la stabilité ni la répétabilité des décisions sur des exécutions multiples.

3.5 CONTEXTE, EXPLICABILITÉ ET RÉPÉTABILITÉ

Un avantage clé des LLMs réside dans leur capacité à générer des justifications textuelles accompagnant leurs décisions. Cette propriété est particulièrement pertinente dans le contexte des opérations SOC, où la confiance dans les systèmes d'assistance automatisés dépend fortement de la capacité à comprendre et à expliquer les décisions prises.

Cependant, l'explicabilité des décisions produites par les LLMs demeure conditionnée à la qualité et à la pertinence de l'information fournie en entrée. En l'absence de contexte organisationnel explicite, les justifications générées peuvent apparaître plausibles tout en reposant sur des hypothèses erronées ou incomplètes. Ce phénomène est étroitement lié au problème des hallucinations, par lequel un modèle génère des réponses cohérentes en apparence mais factuellement ou contextuellement incorrectes.

La nature probabiliste des LLMs implique également qu'à paramètres et entrées identiques, un même modèle peut produire des sorties différentes d'une exécution à l'autre. Cette variabilité soulève des enjeux majeurs de fiabilité dans un environnement critique tel qu'un SOC, où une même alerte pourrait être classifiée différemment selon l'exécution du modèle.

La notion de répétabilité, définie comme la capacité d'un modèle à produire des sorties de signification cohérente lorsqu'un même cas est évalué à plusieurs reprises dans des conditions strictement identiques, constitue ainsi un critère essentiel pour une intégration opérationnelle des LLMs Shyr *et al.* (2025).

Sans une stabilité minimale des décisions, l'automatisation ou l'assistance au triage des alertes risque d'introduire une nouvelle source d'incertitude plutôt que de réduire la charge cognitive des analystes. Ces observations renforcent l'idée que le contexte et l'évaluation de la répétabilité ne doivent pas être considérés comme des aspects secondaires, mais comme des conditions nécessaires à l'utilisation fiable des LLMs dans les environnements SOC.

3.6 SYNTHÈSE

Ce chapitre a présenté les principes fondamentaux des modèles de langage de grande taille, leurs applications potentielles en cybersécurité et leurs limites spécifiques dans le cadre du triage des alertes SOC. Il a mis en évidence que, malgré leurs capacités avancées de traitement du langage et de génération d'explications, les LLMs restent fortement dépendants de l'information contextuelle disponible pour produire des décisions fiables, explicables et stables.

Ces constats confirment la nécessité d'étudier de manière systématique l'impact du contexte organisationnel sur les décisions de triage produites par les LLMs, tout en tenant compte de la répétabilité des classifications dans des conditions contrôlées. Le chapitre suivant s'appuie sur cette analyse pour décrire l'environnement expérimental, la méthodologie et l'architecture du système mis en œuvre afin d'évaluer empiriquement cette influence dans un cadre SOC contrôlé.

CHAPITRE IV
ENVIRONNEMENT EXPÉRIMENTAL ET MÉTHODOLOGIE

Ce chapitre présente l'approche expérimentale et la méthodologie mises en œuvre afin d'évaluer l'impact de l'intégration du contexte organisationnel sur la précision de la classification des alertes fondée sur LLMs. L'évaluation est réalisée dans un environnement de type cyber range qui combine l'émulation d'attaques et la collecte centralisée des journaux d'événements.

4.0.1 ENVIRONNEMENT EXPÉRIMENTAL

L'expérience a été réalisée au sein du **Cyber-Québec cyber range**, un environnement virtualisé de formation et de recherche en cybersécurité, construit sur **OpenNebula** pour l'orchestration des machines virtuelles et **FireEdge** pour un accès distant sécurisé.

CyberQuébec permet le déploiement de topologies réseau complexes, de scénarios de simulation d'attaques et d'infrastructures de surveillance défensive. Cet environnement contrôlé offre des conditions SOC réalistes mais isolées, garantissant la reproductibilité sans impact sur les systèmes de production. L'environnement comprend quatre machines virtuelles (VM) interconnectées, reproduisant une architecture SOC typique :

— **Serveur d'émulation adverse (CALDERA)** : Une VM Ubuntu 22.04 hébergeant la plateforme MITRE CALDERA, utilisée pour simuler des tactiques, techniques et procédures (TTP) adverses.

— **Poste de travail Linux surveillé** : Un endpoint Ubuntu 22.04 équipé des agents Wazuh et Sandcat pour la surveillance des activités.

— **Poste de travail Windows surveillé** : Un endpoint Windows 11 configuré de manière similaire, exécutant les agents Wazuh et CALDERA.

— **Serveur de supervision (Wazuh)** : Une machine virtuelle Ubuntu 22.04 dédiée à la centralisation et à la normalisation des journaux via le gestionnaire et le tableau de bord Wazuh.

Les journaux générés par les deux endpoints sont collectés et normalisés par Wazuh avant d'être exportés au format JSON. Ces données sont ensuite transférées vers un poste de travail Windows pour être traitées et analysées.

Cette architecture intégrée permet d'évaluer le comportement du modèle dans un environnement simulé représentatif des opérations SOC Oniagbi (2024); Vielberth *et al.* (2020); Kersten *et al.* (2023). La Figure 4.1 illustre la configuration expérimentale, en mettant en évidence les interactions entre la plateforme d'émulation d'adversaires CALDERA, le SIEM Wazuh et le modèle Meta LLaMA 3.

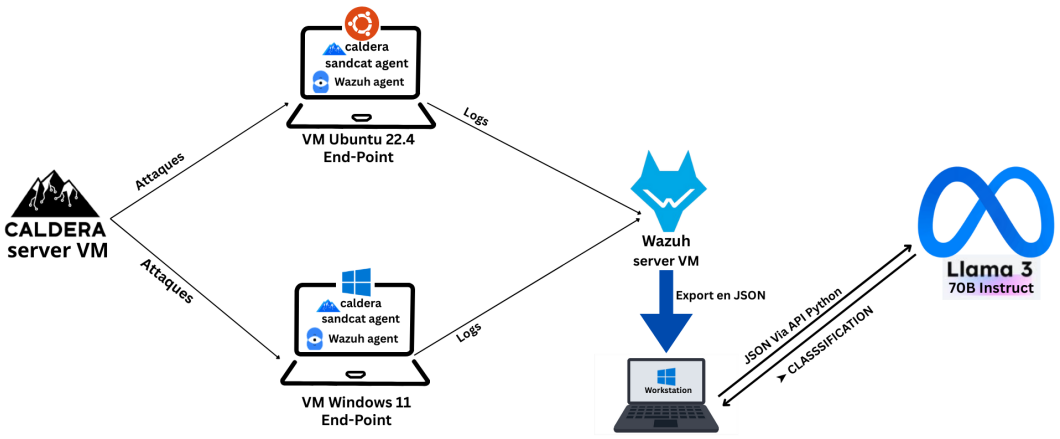

FIGURE 4.1 : Architecture expérimentale intégrant CALDERA, les endpoints Windows et Linux surveillés par Wazuh, ainsi que le modèle LLaMA 3.3 pour la classification des alertes

4.0.2 GÉNÉRATION ET ÉTIQUETAGE DU JEU DE DONNÉES

En suivant l'approche de Ban et al. Ban *et al.* (2021), la collecte et la génération des alertes ont été réalisées à l'aide de Wazuh, configuré pour recevoir les événements provenant des deux hôtes surveillés. Afin de produire des alertes pertinentes, la plateforme CALDERA a été utilisée pour déclencher plusieurs scénarios d'attaque basés sur la matrice MITRE ATT &CK. Ces scénarios ont permis d'émuler diverses tactiques, notamment l'exécution de scripts PowerShell (T1059), le transfert d'outils malveillants (T1105), l'élévation de privilèges (T1548) et la création de comptes non autorisés (T1136).

Les figures 4.2 et 4.3 présentent respectivement les tableaux de bord de Wazuh et de CALDERA, illustrant l'infrastructure de supervision ainsi que l'environnement d'émulation adversaire utilisés pour générer les alertes analysées dans cette étude.

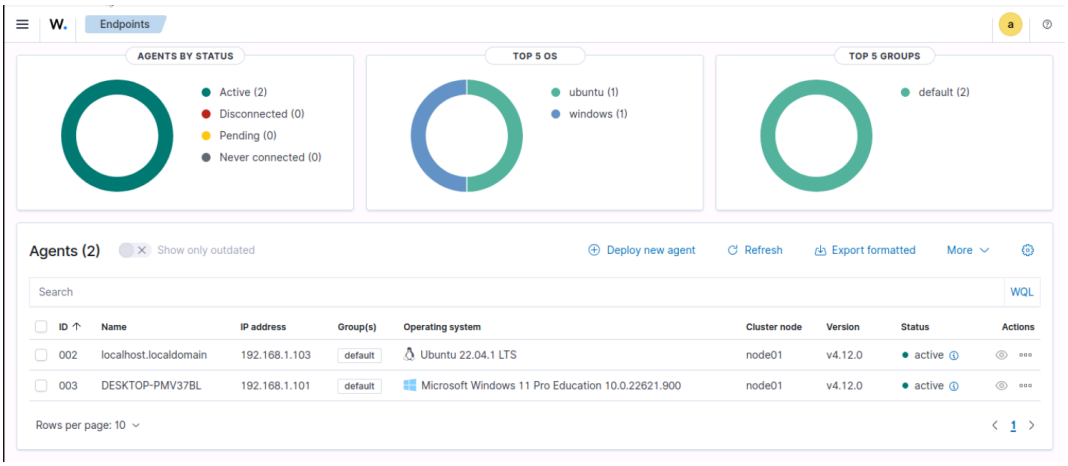

FIGURE 4.2 : Page d'accueil de Wazuh avec les deux agents connectés visibles

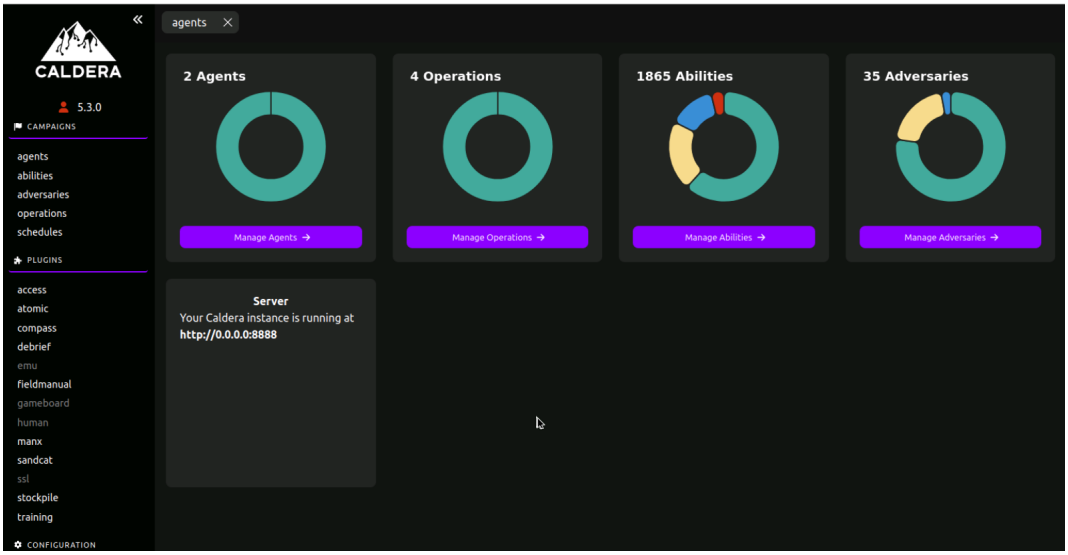

FIGURE 4.3 : Page d'accueil Caldera avec les deux agents connectés visibles

Ces attaques ont été menées contre des endpoints Linux et Windows, déclenchant la génération d'alertes par le moteur Wazuh. En parallèle, des activités légitimes telles que des authentifications valides, des exécutions administratives, des mises à jour logicielles et de la navigation dans les répertoires ont été réalisées afin de produire des événements non malveillants. L'ensemble des journaux collectés a été exporté depuis Wazuh au format JSON

pour analyse. Le jeu de données final comprend 55 alertes distinctes, dont 21 provenant du poste Linux et 34 de Windows, comme résumé dans le Tableau 4.1. Le jeu de données est équilibré entre les alertes intéressantes **(Interesting)** correspondant à des activités d'attaque et les alertes non intéressantes (**Not Interesting**) correspondant à des opérations normales. Chaque alerte a été étiquetée manuellement sur la base de la connaissance du scénario d'émulation et du contexte organisationnel, conformément à la méthodologie de triage cognitif décrite par Zhong et al. Zhong *et al.* (2016). Une alerte SIEM complète générée lors de l'expérimentation est présentée en annexe A afin d'illustrer la structure des données analysées par le modèle.

TABLEAU 4.1 : Répartition de l'ensemble de données par système d'exploitation et type d'alerte

Système d'exploitation	**Interesting Alerts**	**Not-interesting Alerts**
Windows 11	18	16
Linux 22.4 LTS	11	10
TOTAL	29	26

4.0.3 INTÉGRATION DU LLM ET DES INFORMATIONS CONTEXTUELLES

Le modèle utilisé dans cette étude est **Meta LLaMA 3.3-70B-Instruct-Turbo**, accessible via l'API Together. Un script Python a été développé pour soumettre des alertes au format JSON au modèle, en le sollicitant pour agir comme un analyste SOC de niveau 1 chargé d'évaluer la pertinence des alertes. Contrairement aux méthodes automatisées de génération de prompts telles qu'AutoPrompt Shin *et al.* (2020), qui reposent sur une recherche algorithmique pour optimiser les jetons déclencheurs, les prompts utilisés dans cette étude ont été conçus manuellement afin de refléter des workflows réalistes de classification d'alertes dans un SOC.

4.0.4 TOGETHER API COMME INTERFACE D'ACCÈS AUX LLMS

Dans les environnements applicatifs modernes, l'accès aux modèles de langage de grande taille s'effectue le plus souvent via des interfaces de programmation applicative (API) fournies par des plateformes spécialisées. Ces plateformes abstraient les contraintes liées à l'hébergement, à la gestion des ressources matérielles et au déploiement des modèles, tout en offrant une interface standardisée pour leur interrogation.

Together API s'inscrit dans cette approche en proposant un accès unifié à plusieurs modèles de langage de grande taille hébergés, via des requêtes structurées de type conversationnel. L'interaction avec le modèle repose sur l'envoi de messages décrivant la tâche à effectuer, et sur la récupération de réponses textuelles générées par le modèle en sortie.

Dans le cadre de ce mémoire, Together API est utilisée comme mécanisme d'accès au modèle de langage, permettant d'illustrer concrètement le processus d'interrogation d'un LLM dans un contexte applicatif. Un exemple simplifié d'appel à l'API est présenté à la Figure 4.4, afin d'illustrer la structure générale des requêtes utilisées.

```python
from together import Together

client = Together() # auth defaults to os.environ.get("TOGETHER_API_KEY")

response = client.chat.completions.create(
    model="meta-llama/Llama-3.3-70B-Instruct-Turbo",
    messages=[
      {
        "role": "user",
        "content": "What are some fun things to do in New York?"
      }
    ]
)
print(response.choices[0].message.content)
```

FIGURE 4.4 : Exemple d'appel à Together API pour l'interrogation d'un LLM

4.0.5 TRIAGE DES ALERTES DE NIVEAU 1 (TIER-1) DANS LES OPÉRATIONS SOC

Le triage des alertes de niveau 1 (Tier-1) correspond à la première couche opérationnelle d'un SOC, où les analystes déterminent rapidement si une alerte est bénigne, correspond à un faux positif ou indique un incident potentiel. Les analystes Tier-1 s'appuient sur une combinaison de métadonnées techniques et de connaissances organisationnelles pour formuler des jugements rapides sous pression temporelle. Cette distinction est pertinente pour notre expérience : la configuration « sans contexte » reproduit uniquement la couche technique de l'analyse Tier-1, tandis que la configuration « avec contexte » se rapproche davantage du processus de classification réel d'un analyste Tier-1 en intégrant des connaissances organisationnelles. Deux scénarios expérimentaux ont été menés :

SCÉNARIO 1 : SANS CONTEXTE ORGANISATIONNEL

Ce premier scénario correspond à une configuration de triage des alertes sans prise en compte du contexte organisationnel. Dans cette configuration de référence, le modèle Meta LLaMA 3.3–70B-Instruct-Turbo reçoit uniquement les attributs techniques bruts extraits des journaux Wazuh, notamment la description de la règle, le type d'événement, l'horodatage, l'utilisateur, l'hôte et le niveau de sévérité.

Aucune information relative à l'organisation surveillée n'est fournie au modèle. En particulier, celui-ci ne dispose d'aucune connaissance concernant les rôles des utilisateurs, les comportements attendus, les plages horaires normales d'activité ou les contraintes opérationnelles. Le classification repose donc exclusivement sur l'interprétation des indicateurs techniques de bas niveau présents dans l'alerte.

Cette configuration reproduit une situation de triage limitée, comparable à celle d'un analyste Tier-1 ne disposant que des métadonnées SIEM, sans accès au contexte organisationnel nécessaire à une interprétation fine des événements. Elle sert de scénario de référence afin d'évaluer l'impact de l'intégration du contexte organisationnel sur les décisions de triage. Le prompt utilisé pour le triage des alertes dans ce scénario est présenté à la figure 4.5.

```
{
 message = (
     f"You are an experienced Tier 1 SOC analyst. Review the SIEM alerts
      ↪  and "
     f"decide which alerts are interesting or not-interesting. "
     f"An interesting alert indicates a potential incident that needs "
     f"to be investigated further. "
     f"A non-interesting alert is either informational or a false positive.
      ↪  "
     f"Make this decision based on the details in the alert and respond
      ↪  with "
     f"your decision. There is one Wazuh alert in JSON format. "
     f"Please review the alert and indicate whether it is 'interesting' "
     f"or 'not-interesting'. Your message should use the following format:
      ↪  "
     f"{alert_id:} <ID> | {alert_description:} <short> | "
     f"{alert_decision:} <interesting|not-interesting> | "
     f"{reason:} <brief>. \n\n"
     f"Alert: {alert}"
 )
}
```

FIGURE 4.5 : Prompt utilisé pour le triage des alertes sans informations contextuelles

SCÉNARIO 2 : AVEC CONTEXTE ORGANISATIONNEL

Dans ce scénario, le triage des alertes est réalisé en fournissant explicitement au modèle un contexte organisationnel. En plus de l'alerte SIEM brute, le modèle reçoit un fichier structuré *context.json* décrivant les éléments clés de l'organisation surveillée. Ces informations contextuelles incluent les rôles et appartenances aux groupes des utilisateurs, les réseaux auto-

risés, les niveaux de privilèges attendus, les heures d'ouverture, les fenêtres de maintenance et les politiques de sécurité générales. L'objectif de cette configuration est de se rapprocher du processus de classification d'un analyste SOC de niveau 1, qui interprète systématiquement les alertes en corrélant les indicateurs techniques aux pratiques et aux contraintes organisationnelles. En injectant directement ces informations contextuelles dans le prompt, le modèle est en mesure d'évaluer si une alerte donnée est cohérente avec le comportement opérationnel normal ou si elle est indicative d'un incident de sécurité potentiel. La structure complète du contexte organisationnel injecté dans les invites du modèle est fournie en Annexe B. Cette contextualisation permet au modèle d'évaluer les caractéristiques techniques des événements en tenant compte de l'environnement opérationnel de l'organisation, reproduisant ainsi le processus de classification utilisé par les analystes Tier-1 lors du triage des alertes.

Le prompt utilisé pour le triage des alertes avec contexte organisationnel est présenté à la Figure 4.6.

```
{
message = (
    f"You are an experienced Tier 1 SOC analyst. Review the following "
    f"SIEM alert and decide if it is 'interesting' or 'not-interesting'. "
    f"An interesting alert indicates a potential incident requiring "
    f"investigation, while a non-interesting alert is informational or
     ↪ benign. "
    f"Use the BUSINESS CONTEXT as a guide to understand the organization's
     ↪ "
    f"normal behavior (its users, groups, assets, business hours, and
     ↪ networks). "
    f"Analyze the alert in this context and make your own judgment. "
    f"In your explanation, clearly mention which factors from the context
     ↪ "
    f"influenced your decision. For example: user group, account type, "
    f"business hours, host role, network location, or any other "
    f"relevant contextual details.\n\n"
    f"Respond strictly using this format:\n"
    f"{alert_id:} <ID> | {alert_description:} <short> | "
    f"{alert_decision:} <interesting|not-interesting> | "
    f"{reason:} <brief explanation mentioning contextual factors>\n\n"
    f"{ALERT:}\n{json.dumps(alert, ensure_ascii=False)}\n\n"
    f"BUSINESS CONTEXT:\n{json.dumps(context, ensure_ascii=False)}"
)
}
```

FIGURE 4.6 : Prompt utilisé pour le triage des alertes avec contexte organisationnel

Chaque décision produite par le modèle devait être accompagnée d'une justification explicite établissant un lien clair entre les données brutes issues des journaux SIEM et les informations contextuelles injectées. Cette exigence vise à contraindre le modèle à expliciter la classification, de manière analogue au processus cognitif suivi par les analystes SOC de niveau Tier-1 lors du triage des alertes.

Cette approche prolonge la méthodologie proposée par Oniagbi Oniagbi (2024), en y intégrant explicitement la dimension contextuelle organisationnelle, dont l'importance a été mise en évidence par Alahmadi et al. Alahmadi *et al.* (2022) ainsi que par Sharma et al. Sharma *et al.* (2025). Contrairement aux travaux existants, où le contexte est soit absent, soit implicitement intégré, le présent travail considère le contexte comme une variable expérimentale contrôlée, injectée de manière structurée dans les entrées du modèle.

Le processus décisionnel de triage des alertes, tant dans le scénario sans contexte que dans le scénario enrichi par le contexte organisationnel, est illustré respectivement aux Figures 4.7 et 4.8.

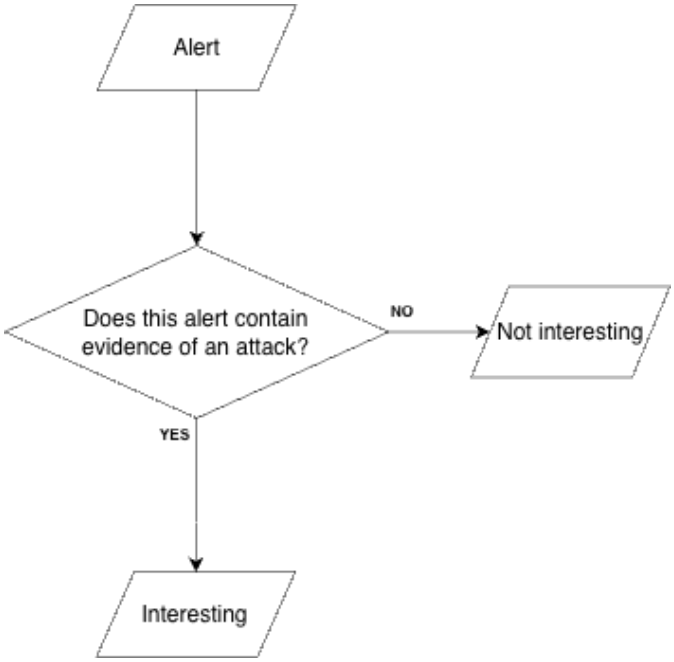

FIGURE 4.7 : Flux de décision du triage des alertes sans contexte organisationnel

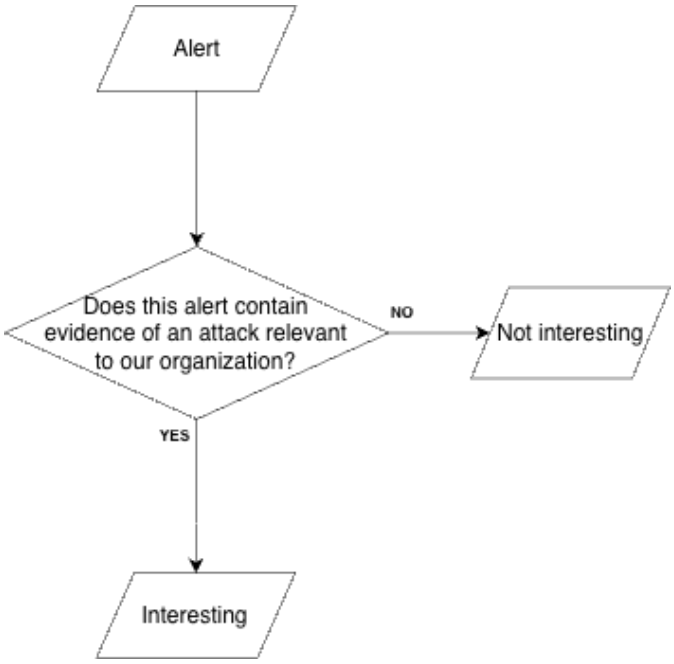

FIGURE 4.8 : Flux de décision du triage des alertes avec contexte organisationnel

4.0.6 MÉTRIQUES D'ÉVALUATION

Les performances du modèle ont été évaluées à l'aide de métriques classiques de classification, à savoir la précision globale (*accuracy*), la valeur prédictive positive (*precision*), le rappel (*recall*), le score F1 et le taux de faux positifs (*false positive rate*, FPR). Ces métriques sont largement reconnues comme fondamentales pour l'évaluation des performances des modèles de classification, notamment dans des contextes opérationnels, comme le soulignent Muntean et Militaru Muntean & Militaru (2023) ainsi qu'Oniagbi Oniagbi (2024).

Les métriques utilisées sont définies comme suit :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

L'*accuracy* mesure la proportion globale d'alertes correctement triées par le modèle sur l'ensemble du jeu de données.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.2}$$

La *precision* évalue la fiabilité des alertes classées comme intéressantes par le modèle, en quantifiant la proportion d'alertes effectivement pertinentes parmi celles signalées comme telles.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.3}$$

Le *recall* mesure la capacité du modèle à identifier l'ensemble des alertes réellement pertinentes, en évaluant la proportion d'alertes intéressantes correctement détectées.

$$F1_{\text{score}} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4.4}$$

Le $F1_{\text{score}}$ correspond à la moyenne harmonique de la précision et du rappel, offrant un compromis entre ces deux métriques lorsque l'équilibre entre faux positifs et faux négatifs est critique.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{4.5}$$

Dans un contexte SOC, le taux de faux positifs constitue un indicateur particulièrement important. Il quantifie la proportion d'alertes bénignes incorrectement signalées comme intéressantes, contribuant directement à la charge de travail des analystes et au phénomène de fatigue liée aux alertes.

Où :

— TP (*True Positives*) correspond aux alertes intéressantes correctement identifiées.

— TN (*True Negatives*) correspond aux alertes bénignes correctement ignorées.

— FP (*False Positives*) correspond aux alertes bénignes incorrectement signalées comme intéressantes.

— FN (*False Negatives*) correspond aux alertes intéressantes non détectées par le modèle.

4.0.7 ÉVALUATION DE LA RÉPÉTABILITÉ

Afin d'évaluer la répétabilité des décisions produites par le modèle, nous avons suivi un protocole inspiré de la définition proposée par Shyr et al. Shyr *et al.* (2025), qui définissent la répétabilité comme la capacité d'un modèle à produire des sorties de signification cohérente lorsqu'un même cas est évalué de manière répétée dans des conditions strictement identiques. L'application de ce principe implique de maintenir constants l'ensemble des facteurs externes afin que toute variation observée puisse être attribuée exclusivement au comportement intrinsèque du modèle.

Dans le cadre de cette expérimentation, chaque alerte a été soumise au modèle 100 fois en utilisant le même LLM (Meta LLaMA 3.3–70B-Instruct-Turbo), le même prompt, la même alerte en entrée avec une structure JSON strictement identique, ainsi que la même configuration de génération. Aucun paramètre de génération, incluant notamment la température, le *top-p*, le *top-k* ou tout autre paramètre, n'a été modifié entre les exécutions. Les paramètres par défaut de l'API Together ont été conservés pour l'ensemble des essais. Cette configuration garantit que l'évaluation isole la variabilité intrinsèque du modèle, sans introduire d'artefacts liés à des changements de configuration.

La répétabilité a été mesurée en calculant, pour chaque alerte i, la proportion d'exécutions produisant la même décision de triage. Si le modèle génère c_i décisions identiques sur un total de $N = 100$ exécutions, le score de répétabilité pour cette alerte est défini comme suit :

$$r_i = \frac{c_i}{N}. \tag{4.6}$$

Un score de répétabilité global est ensuite obtenu en moyennant les valeurs r_i sur l'ensemble du jeu de données, composé de $M = 55$ alertes :

$$R_{\text{global}} = \frac{1}{M} \sum_{i=1}^{M} r_i. \tag{4.7}$$

Une valeur de $R_{\text{global}} = 1$ indique une cohérence complète des décisions de triage pour toutes les alertes sur l'ensemble des exécutions, tandis que des valeurs inférieures traduisent une variabilité des classifications produites par le modèle. Bien que la métrique proposée par Shyr et al. Shyr *et al.* (2025) ait été initialement conçue pour l'évaluation de sorties textuelles longues en contexte clinique, le principe sous-jacent de cohérence sous conditions identiques s'adapte naturellement à notre cadre expérimental, dans lequel l'accord peut être évalué au niveau des étiquettes de classification binaire.

CHAPITRE V
RÉSULTATS EXPÉRIMENTAUX

Ce chapitre présente les résultats obtenus lors des expérimentations menées avec et sans contexte organisationnel. Il décrit le jeu de données utilisé, compare les performances quantitatives du modèle, analyse la répétabilité des décisions et illustre qualitativement l'effet du contexte sur le triage des alertes SOC.

5.1 VUE D'ENSEMBLE DES EXPÉRIENCES

Les expériences ont été menées selon les deux configurations définies au Chapitre 4 : (i) un scénario de référence sans contexte organisationnel, dans lequel le modèle traite uniquement les métadonnées brutes des alertes SIEM, et (ii) un scénario contextualisé, dans lequel les métadonnées des alertes sont enrichies par un contexte organisationnel structuré fourni via un fichier JSON externe.

L'objectif de ce chapitre est de présenter et de comparer les résultats obtenus dans ces deux scénarios à l'aide de métriques quantitatives de performance et de mesures de répétabilité. L'interprétation de ces résultats et leurs implications pour les opérations SOC sont discutées au Chapitre 6.

5.2 DESCRIPTION DU JEU DE DONNÉES ET VÉRITÉ TERRAIN

L'évaluation a été réalisée sur un jeu de données composé de 55 alertes SIEM générées par la plateforme Wazuh à partir de deux endpoints surveillés : un poste Windows 11 et un système Linux Ubuntu 22.04. Parmi ces alertes, 29 ont été étiquetées comme intéressantes, correspondant à des activités malveillantes ou suspectes générées par des scénarios d'émulation

d'adversaires exécutés avec CALDERA. Les 26 alertes restantes ont été étiquetées comme non intéressantes, correspondant à des comportements système bénins, à des activités utilisateur légitimes et à des opérations administratives routinières.

Les étiquettes de vérité terrain ont été attribuées manuellement sur la base d'une connaissance complète des scénarios d'attaque exécutés et de l'environnement organisationnel, en suivant les principes de triage décrits par Zhong et al. Zhong *et al.* (2016). Toutes les alertes ont été exportées au format JSON afin de préserver l'ensemble des métadonnées nécessaires à la classification par le LLM. La Figure 5.1 illustre la distribution des alertes en fonction des systèmes d'exploitation et des étiquettes de vérité terrain.

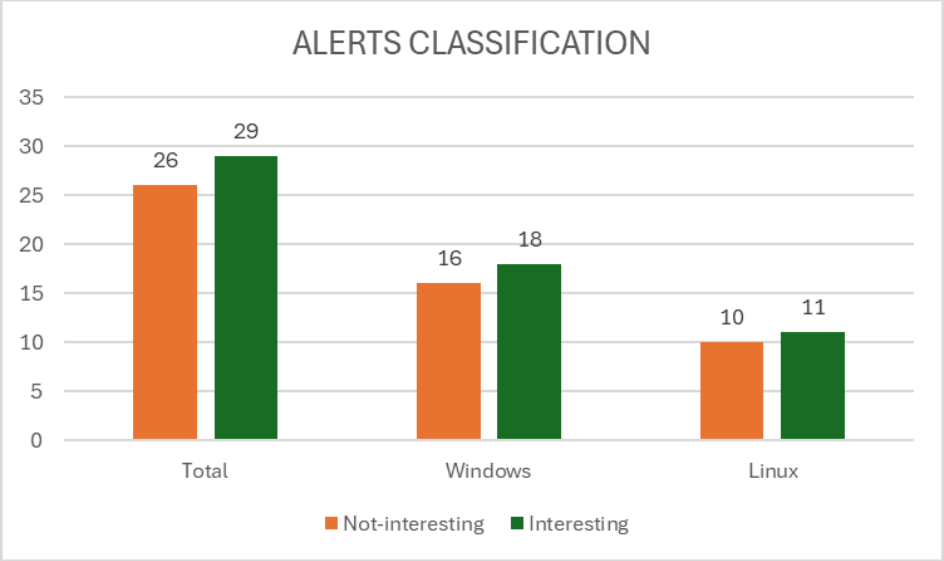

FIGURE 5.1 : Composition du jeu de données

Les Tableaux 5.1 et 5.2 résument respectivement les alertes générées sur les systèmes Windows et Linux, ainsi que les étiquettes de vérité terrain attribuées manuellement.

Afin de ne pas alourdir le corps du mémoire, les tableaux complets, incluant pour chaque alerte l'identifiant, la description, l'identifiant MITRE ATT&CK, la technique, la tactique et l'étiquette de vérité terrain, sont fournis à l'Appendice C.

TABLEAU 5.1 : Vérité terrain des alertes Windows

Alerte ID	Description	Ground Truth
1757612118.1653432	Powershell process spawned Windows command shell instance	Interesting
1757612081.1635793	Powershell process created an executable file in Windows root folder	Interesting
1757612080.1632975	Executable file dropped in folder commonly used by malware	Interesting
1757612033.1626234	Windows command prompt started by an abnormal process	Interesting
1757611970.1617034	Powershell was used to delete files or directories	Interesting
1757611774.1589886	Value added to registry key has Base64-like pattern	Interesting
1757611590.1558176	powershell.exe created a new scripting file under Windows Temp or User data folder	Interesting
1757611525.1549468	Sysmon - Suspicious Process - lsm.exe	Interesting
1757613944.1876416	Registry Value Entry Added to the System	Interesting
1757611467.1517244	Command shell started script with /c modifier	Interesting
1757611465.1409874	Executable dropped in Windows root folder	Interesting
1757611399.1392111	sdclt.exe launched via powershell.exe with high integrity level	Interesting
1757611185.1319274	Powershell process invoked auto-elevated utility ComputerDefaults.exe (possible UAC bypass)	Interesting
1757611185.1335881	Command interpreter added to registry key associated to UAC bypass by auto-elevated processes	Interesting
1757611047.1222736	Powershell spawned a powershell process which executed a base64 encoded command	Interesting
1757610816.1173186	Powershell script compiling code using CSC.exe (possible malware drop)	Interesting
1757610654.1112794	Sysmon - Suspicious Process - lsass	Interesting
1757610608.1101539	Executable file dropped in Users/Public folder	Interesting
1757607634.840437	Executable file created in AppData temp folder	Interesting
1757657178.471270	Registry Value Entry Deleted	Interesting
1757946519.862756	Special privileges assigned to new logon	Interesting
1757612453.1671822	Windows Logon Success	Not Interesting
1757613591.1821187	Service startup type was changed	Not Interesting
1757612118.1647649	Suspicious Windows cmd shell execution	Not Interesting
1757613583.1810657	Software protection service scheduled successfully	Not Interesting
1757605861.812840	License activation (slui.exe) failed	Not Interesting
1757613920.1840104	Registry Key Integrity Checksum Changed	Not Interesting
1757656769.388902	A net.exe account discovery command was initiated	Not Interesting
1757946519.877972	Non service account logged off	Not Interesting
1757946519.865952	Windows Workstation Logon Success	Not Interesting
1757946508.845453	Logon Failure - Unknown user or bad password	Not Interesting
1757945940.790942	Software protection service scheduled successfully	Not Interesting
1757915969.377664	Discovery activity executed	Not Interesting
1757797573.1708146	The VSS service is shutting down due to idle timeout	Not Interesting

TABLEAU 5.2 : Vérité terrain des alertes Linux

Alerte ID	Description	Ground Truth
1757448450.1607664	User changed password	Not Interesting
1757421591.908968	PAM : Login session closed	Not Interesting
1757421591.908504	PAM Login session opened	Not Interesting
1761280350.215525	syslog : User authentication failure	Not Interesting
1761280233.214804	Wazuh agent started	Not Interesting
1757439593.1276633	Integrity checksum changed	Interesting
1757439233.1227872	CIS Ubuntu Linux 22.04 LTS Benchmark v2.0.0. : Ensure a nftables table exists. : Status changed from failed to passed	Not Interesting
1757396591.314337	New group added to the system	Interesting
1757390403.187211	Apparmor DENIED	Not Interesting
1757439593.1274699	File added to /etc/cron.d/persistevil	Interesting
1757302739.245548	Systemd : Service exited due to a failure.	Not Interesting
1757448450.1606656	Successful sudo to ROOT executed.	Not Interesting
1757448450.1605639	New user added to the system	Interesting
1757439237.1234145	SCA summary : CIS Ubuntu Linux 22.04 LTS Benchmark v2.0.0. : Score less than 50% (44)	Interesting
1759427649.2846606	PAM : User login failed.	Not Interesting
1757072746.14293	The CVE-2023-28755 that affected ruby-rubygems was solved due to an update in the agent or feed.	Not Interesting
1757069230.9758	New dpkg (Debian Package) installed	Not Interesting
1757556911.71617	Failed attempt to run sudo	Interesting
1756921222.1803596	CIS Ubuntu Linux 22.04 LTS Benchmark v2.0.0. : Ensure permissions on /etc/shells are configured.	Interesting
1756921222.1806864	CIS Ubuntu Linux 22.04 LTS Benchmark v2.0.0. : Ensure permissions on /etc/security/opasswd are configured.	Not Interesting
1758964101.591531	File deletion	Not Interesting

5.3 RÉSULTATS QUANTITATIFS SANS CONTEXTE ORGANISATIONNEL

Les performances quantitatives du modèle en l'absence de contexte organisationnel sont résumées dans le Tableau 5.3.

TABLEAU 5.3 : Résultats de l'analyse sans contexte

Metric	Value
TP (True Positive)	28
TN (True Negative)	19
FP (False Positive)	7
FN (False Negative)	1
Total	55
Precision	0,80
Accuracy	0,8545
Recall	0,9655
F1 Score	0,875

Dans cette configuration de référence, le modèle atteint une précision globale (accuracy) de 85,45%, avec une précision (precision) de 0,80 et un rappel (recall) de 0,97. Sur les 29 alertes intéressantes, 28 sont correctement identifiées (TP = 28), tandis qu'une alerte malveillante est manquée (FN = 1). Toutefois, le modèle génère sept faux positifs (FP = 7), en classant à tort des alertes bénignes comme intéressantes.

Le score F1 de 0,875 reflète un compromis entre une forte sensibilité et une précision réduite. Ces résultats indiquent qu'en l'absence de contexte organisationnel, le modèle adopte une stratégie de triage conservatrice privilégiant la détection, au prix d'une augmentation de la charge de travail des analystes.

5.4 RÉSULTATS QUANTITATIFS AVEC CONTEXTE ORGANISATIONNEL

Lorsque le contexte organisationnel est intégré au processus de triage des alertes, les performances du modèle s'améliorent de manière significative. Le Tableau 5.4 présente les résultats quantitatifs obtenus dans cette configuration.

TABLEAU 5.4 : Résultats de l'analyse avec contexte

Metric	Value
TP (True Positive)	29
TN (True Negative)	26
FP (False Positive)	0
FN (False Negative)	0
Total	55
Precision	1,00
Accuracy	1,00
Recall	1,00
F1 Score	1,00

Dans ce scénario, le modèle classe correctement l'ensemble des 55 alertes, atteignant une valeur de 1,00 pour toutes les métriques évaluées, y compris l'accuracy, la précision, le rappel et le score F1 (tous égaux à 1,00). Aucun faux positif ni faux négatif n'est observé (FP = 0, FN = 0).

Ces résultats montrent que la disponibilité d'un contexte organisationnel structuré permet au modèle de résoudre les ambiguïtés présentes dans les alertes SIEM brutes et d'aligner ses décisions sur les comportements opérationnels attendus.

5.5 ANALYSE COMPARATIVE DES PERFORMANCES

Une comparaison directe entre les deux scénarios expérimentaux met en évidence l'impact du contexte organisationnel sur les performances du triage des alertes. La Figure 5.2 présente une comparaison visuelle des performances du modèle avec et sans information contextuelle.

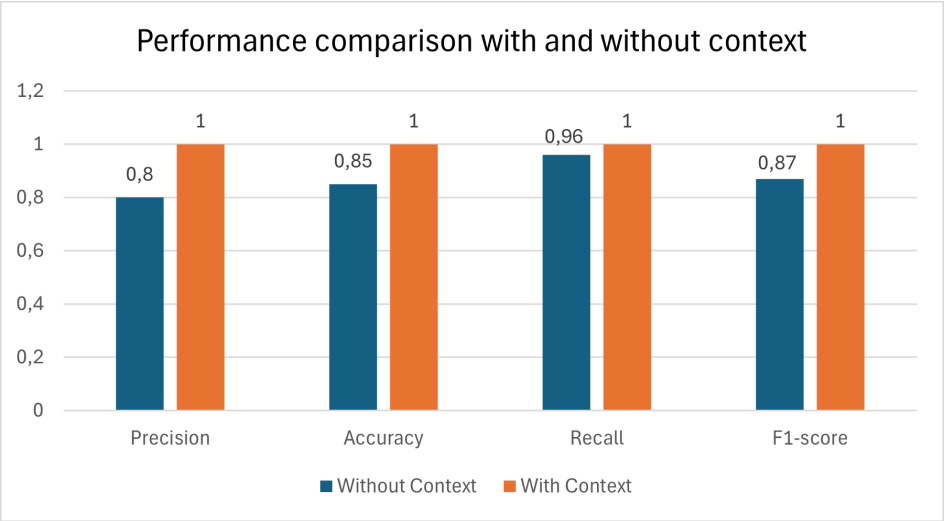

FIGURE 5.2 : Comparaison des performances avec et sans contexte organisationnel

L'amélioration la plus marquée concerne la réduction des faux positifs, qui passent d'un niveau non négligeable dans le scénario sans contexte à zéro lorsque le contexte est fourni. Bien que le rappel demeure élevé dans les deux configurations, l'intégration du contexte organisationnel entraîne une augmentation notable de la précision et de la fiabilité globale des décisions.

5.6 RÉPÉTABILITÉ DES DÉCISIONS DU MODÈLE

Afin d'évaluer la stabilité des décisions, chaque alerte a été soumise 100 fois dans des conditions strictement identiques dans les deux scénarios expérimentaux. En présence du contexte organisationnel, le modèle produit des décisions cohérentes pour toutes les exécutions, conduisant à un score global de répétabilité :

$$R_{\text{global}} = 1. \tag{5.1}$$

Sans contexte organisationnel, la répétabilité globale demeure élevée mais présente de légères variations pour un petit sous-ensemble d'alertes, aboutissant à un score global strictement

inférieur à 1 ($R_{\text{global}} < 1$). Ces variations indiquent qu'en l'absence d'ancrage contextuel, le processus de classification interne du modèle peut fluctuer lorsqu'il traite des alertes ambiguës. L'intégration du contexte organisationnel élimine cette variabilité et conduit à des décisions de triage Tier-1 stables et répétables.

5.7 ANALYSE QUALITATIVE SANS CONTEXTE ORGANISATIONNEL

Conformément aux observations rapportées dans la littérature Alahmadi *et al.* (2022), l'absence de contexte organisationnel induit un biais de prudence dans le comportement du modèle. Face à des alertes ambiguës, le modèle tend à classer les événements comme intéressants, ce qui entraîne des faux positifs et, dans certains cas, des incohérences entre exécutions répétées.

Par exemple, l'alerte « Logon Failure – Unknown user or bad password » est interprétée de manière générique comme une tentative potentielle de force brute lorsque le contexte est absent, malgré l'absence d'indices corroborants. La Figure 5.3 présente la décision produite par le modèle pour cette alerte dans la configuration sans contexte organisationnel.

```
{
  "alert_id": "1757946508.845453",
  "alert_description": "Logon Failure - Unknown user or bad password",
  "alert_decision": "interesting",
  "reason": "The alert indicates a logon failure due to an unknown user or
  ↪  bad password, which could be a potential security incident, such as
  ↪  a brute-force attack or unauthorized access attempt."
}
```

FIGURE 5.3 : Classification du modèle sans contexte

La Figure **??** présente un exemple similaire observé sous Linux pour l'alerte « PAM : User changed password ». En l'absence d'informations sur les rôles utilisateurs ou les poli-

tiques organisationnelles, le modèle adopte une interprétation de sécurité générique et signale l'événement comme nécessitant une investigation.

```
{
  "alert_id": "1757448450.1607664",
  "alert_description": "PAM: User changed password",
  "alert_decision": "interesting",
  "reason": "The alert indicates a password change for a user (ARTUser),
  ↪ which could be a legitimate action. However, given the potential
  ↪ security implications, such as unauthorized access or privilege
  ↪ escalation, it warrants further investigation to determine whether
  ↪ the change was authorized."
}
```

FIGURE 5.4 : Classification du modèle sans contexte pour une alerte de changement de mot de passe

Globalement, dans la configuration sans contexte, le modèle parvient à détecter la quasi-totalité des alertes malveillantes, mais se montre excessivement sensible, générant des faux positifs et une stabilité décisionnelle réduite.

5.8 ANALYSE QUALITATIVE AVEC CONTEXTE ORGANISATIONNEL

Lorsque le contexte organisationnel est fourni, le processus de classification du modèle évolue de manière substantielle. Plutôt que de s'appuyer uniquement sur des indicateurs techniques, le modèle intègre des informations relatives aux rôles des utilisateurs, aux groupes d'appartenance, aux plages horaires, aux réseaux autorisés et aux politiques organisationnelles.

Cette contextualisation élimine les faux positifs observés dans le scénario sans contexte et conduit à des classifications cohérentes sur l'ensemble des exécutions répétées. Pour l'alerte « Logon Failure – Unknown user or bad password », le modèle identifie correctement l'événement comme bénin en se référant au rôle du compte, à la machine concernée et à la politique organisationnelle autorisant des échecs d'authentification isolés. La Figure 5.5 présente la

décision produite par le modèle pour cette même alerte lorsque le contexte organisationnel est fourni.

```
{
  "alert_id": "1757946508.845453",
  "alert_description": "Logon Failure - Unknown user or bad password",
  "alert_decision": "not-interesting",
  "reason": "The alert is for a logon failure due to an unknown user or
   ↪ bad password for the account \"env-admin\" on the host
   ↪ \"DESKTOP-MDNNEJF\", which is a windows_workstation. Given the
   ↪ business context, \"env-admin\" is a member of the IT group and has
   ↪ roles including linux_admin and windows_admin, making it a technical
   ↪ account expected to have access to this type of asset. The logon
   ↪ type (2) is also typical for a windows_workstation. Additionally,
   ↪ the policies section states that a single authentication failure is
   ↪ considered not-interesting unless it meets specific thresholds or
   ↪ exceptions, which this event does not appear to meet based on the
   ↪ provided information."
}
```

FIGURE 5.5 : Classification du modèle avec contexte organisationnel

Sous Linux, l'alerte « PAM : User changed password » est évaluée à l'aide des politiques spécifiques au groupe concerné et du contexte temporel. La Figure 5.6 montre que le modèle justifie sa décision en tenant compte du groupe d'appartenance de l'utilisateur, de l'heure de l'événement et des politiques organisationnelles associées.

```
{
  "alert_id": "1757448450.1607664",
  "alert_description": "PAM: User changed password",
  "alert_decision": "interesting",
  "reason": "The alert is interesting because the password change occurred
   ↪ for the user \"ARTUser\", who is part of the \"FINANCE\" group, and
   ↪ password changes for this group require verification, as indicated
   ↪ in the \"facts\" section of the business context, specifically under
   ↪ \"password_change\" and \"requires_verification_for_groups\".
   ↪ Additionally, the change happened outside of standard business hours
   ↪ (20:07:29 is outside the 08:30-17:30 window), which could indicate
   ↪ unusual activity. The host involved, \"localhost.localdomain\", is a
   ↪ linux workstation with expected technical accounts and service
   ↪ processes, but the timing and the group membership of the user whose
   ↪ password was changed warrant further investigation."
}
```


FIGURE 5.6 : Classification du modèle avec contexte pour l'alerte de changement de mot de passe

Ces exemples illustrent comment l'information contextuelle fait évoluer le modèle d'une stratégie de détection générique vers un processus analytique structuré, proche du processus de classification des analystes SOC de niveau 1 (Tier-1).

5.9 SYNTHÈSE DES RÉSULTATS

Les résultats expérimentaux mettent en évidence trois observations majeures. Premièrement, un rappel élevé peut être atteint sans contexte organisationnel, mais au prix d'un nombre accru de faux positifs et d'une stabilité décisionnelle réduite. Deuxièmement, le contexte organisationnel améliore significativement la précision, en éliminant les faux positifs dans le jeu de données évalué. Troisièmement, l'information contextuelle renforce la répétabilité des décisions, stabilisant le comportement du modèle sur des exécutions répétées.

Pris dans leur ensemble, ces résultats démontrent que le contexte organisationnel constitue un facteur clé pour un triage des alertes Tier-1 basé sur des LLMs, fiable, cohérent et opérationnellement pertinent dans les environnements SOC.

CHAPITRE VI

CONCLUSION ET PERSPECTIVES DE RECHERCHE

Ce chapitre présente une synthèse des principales contributions du mémoire, discute les apports méthodologiques et pratiques de l'étude, présente ses limites et propose plusieurs perspectives de recherche futures pour l'intégration des LLMs contextualisés dans les opérations SOC.

6.1 SYNTHÈSE DES CONTRIBUTIONS

Ce mémoire avait pour objectif principal d'évaluer l'impact de l'intégration explicite du contexte organisationnel sur le triage des alertes SOC à l'aide de modèles de langage de grande taille (Large Language Models, LLMs). Contrairement à de nombreux travaux existants qui se concentrent sur la comparaison de modèles ou sur l'optimisation des performances globales, cette étude a cherché à isoler le rôle du contexte en tant que variable expérimentale explicite, dans un environnement SOC contrôlé.

Cette recherche a également donné lieu à un article scientifique publié par IEEE dans les actes de la conférence *2026 IEEE 5th International Conference on AI in Cybersecurity* (ICAIC) Roy & Balde (2026).

Les résultats obtenus montrent que, bien que le modèle Meta LLaMA 3.3–70B-Instruct-Turbo soit capable d'identifier la majorité des alertes nécessitant une investigation à partir des seules métadonnées du SIEM, son comportement en l'absence de contexte organisationnel demeure conservateur et imparfait. Cette configuration entraîne une augmentation du nombre de faux positifs, une diminution de la précision ainsi qu'une variabilité des classifications au cours d'exécutions répétées. Ces observations confirment les limites des approches de triage

fondées uniquement sur des indicateurs techniques de bas niveau, largement documentées dans la littérature SOC.

À l'inverse, l'intégration d'un contexte organisationnel structuré, fourni sous la forme d'un fichier JSON externe, améliore le comportement du modèle. Dans cette configuration, le LLM est capable de relier les événements techniques à des éléments tels que les rôles des utilisateurs, les groupes fonctionnels, les plages horaires, les politiques internes et les contraintes opérationnelles. Cette contextualisation permet de réduire les faux positifs observés dans le scénario sans contexte, d'améliorer la précision globale et de produire des classifications stables au cours des exécutions répétées.

Ces résultats mettent en évidence que les limites observées dans les approches LLM appliquées au triage SOC ne sont pas uniquement liées aux capacités intrinsèques des modèles, mais également à la manière dont l'information contextuelle est représentée, structurée et intégrée dans le processus de triage.

6.2 APPORTS MÉTHODOLOGIQUES ET PRATIQUES

Au-delà des résultats quantitatifs, ce travail apporte plusieurs contributions méthodologiques et pratiques aux recherches sur l'automatisation du triage SOC.

Premièrement, il propose une méthodologie expérimentale reproductible permettant d'évaluer l'impact du contexte organisationnel sur le comportement décisionnel d'un LLM, en combinant des métriques classiques de classification et une évaluation explicite de la répétabilité. Cette approche complète les évaluations ponctuelles couramment utilisées dans les travaux existants et met en évidence la stabilité des décisions comme critère clé pour une adoption opérationnelle en SOC.

Deuxièmement, cette étude démontre l'intérêt de découpler la logique de détection (règles SIEM) de la logique de classification contextuel. Plutôt que d'intégrer le contexte directement dans les règles de détection, ce qui augmente la complexité, la rigidité et les coûts de maintenance, le contexte est ici fourni comme une source externe indépendante. Cette séparation favorise la modularité du système, facilite la mise à jour du contexte organisationnel et améliore la scalabilité de la solution.

Troisièmement, les résultats montrent que les LLMs, lorsqu'ils sont correctement contextualisés, peuvent adopter un processus de classification proche de celui des analystes SOC Tier-1. Le modèle ne se limite plus à une détection générique fondée sur des signatures, mais applique une analyse situationnelle intégrant des éléments organisationnels, ce qui correspond davantage aux pratiques opérationnelles réelles des SOC.

6.3 LIMITES DE L'ÉTUDE

Malgré les résultats encourageants obtenus, ce travail présente plusieurs limites qu'il convient de souligner.

Tout d'abord, l'évaluation a été réalisée sur un jeu de données de taille limitée (55 alertes), généré dans un environnement contrôlé de type cyber range. Bien que cette configuration permette un contrôle précis des scénarios et une annotation fiable de la vérité terrain, elle ne reflète pas toute la diversité, la complexité et le bruit présents dans les environnements SOC opérationnels à grande échelle.

Ensuite, le contexte organisationnel utilisé dans cette étude est statique et fourni sous la forme d'un fichier JSON construit manuellement. Cette représentation simplifiée ne capture pas la dynamique réelle des organisations, où les rôles, les actifs, les politiques et les comportements évoluent en continu. De plus, certaines dimensions du contexte, telles que les

dépendances applicatives complexes, les flux métiers ou les historiques d'incidents ne sont pas prises en compte dans le cadre expérimental actuel.

Enfin, l'étude s'est concentrée sur un seul modèle LLM et un seul schéma de prompt, afin d'isoler l'effet du contexte. Les résultats ne peuvent donc pas être généralisés directement à l'ensemble des architectures de modèles ou des stratégies de prompting existantes, même si les tendances observées sont cohérentes avec la littérature récente.

6.4 PERSPECTIVES DE RECHERCHE FUTURES

Plusieurs axes de recherche peuvent être envisagés pour prolonger et enrichir ce travail.

Un premier axe consiste à évaluer l'approche proposée sur des jeux de données plus larges et plus diversifiés, incluant des alertes issues de SOC réels. Une telle évaluation permettrait d'analyser la robustesse du modèle face au bruit, aux alertes corrélées et aux scénarios d'attaque plus complexes.

Un second axe concerne l'évolution du contexte organisationnel vers une représentation dynamique et automatisée. L'intégration de sources telles que les bases de données de gestion de configuration (CMDB), les systèmes de gestion des identités et des accès (IAM), les outils de gouvernance, de gestion des risques et de conformité (GRC), ainsi que les politiques de sécurité documentées, permettrait de maintenir un contexte à jour sans intervention manuelle. Cette approche ouvrirait la voie à des systèmes de triage contextuel continus et adaptatifs.

Un troisième axe de recherche porte sur l'explicabilité et la confiance. Bien que les LLMs puissent produire des justifications textuelles, il reste nécessaire d'évaluer la qualité, la cohérence et la vérifiabilité de ces explications dans un cadre SOC. Des mécanismes

d'IA explicable adaptés aux LLMs pourraient renforcer la confiance des analystes et faciliter l'intégration de ces outils dans des environnements opérationnels.

Enfin, l'intégration de LLMs contextualisés dans des workflows SOC existants constitue une perspective importante. Plutôt qu'un remplacement des analystes humains, ces modèles pourraient agir comme des assistants de triage Tier-1, réduisant la charge cognitive, limitant la fatigue liée aux alertes et permettant aux analystes de se concentrer sur des tâches à plus forte valeur ajoutée.

CONCLUSION

En conclusion, ce mémoire met en évidence le potentiel des LLMs pour assister le triage des alertes SOC, tout en montrant que leur efficacité dépend fortement de la qualité du contexte organisationnel fourni. Dans la configuration sans contexte, le modèle Meta LLaMA 3.3–70B-Instruct-Turbo parvient à classer correctement la majorité des alertes, mais génère encore des faux positifs lorsqu'il ne dispose pas des informations organisationnelles nécessaires pour distinguer certains événements bénins d'activités suspectes.

L'intégration d'un contexte organisationnel structuré modifie significativement ce comportement. Dans cette configuration, aucun faux positif n'est observé dans le corpus évalué, et les classifications demeurent stables au cours des exécutions répétées réalisées dans les mêmes conditions expérimentales. Ces observations montrent que le contexte ne constitue pas seulement un complément d'information, mais un élément central à une classification plus fiable des alertes SOC par les LLMs.

Ainsi, ce mémoire montre que l'amélioration du triage SOC basé sur l'IA ne repose pas uniquement sur l'utilisation de modèles plus puissants, mais également sur une meilleure représentation, structuration et intégration du contexte organisationnel. L'exploitation de ce contexte apparaît donc comme une condition essentielle pour réduire les faux positifs, renforcer la stabilité des classifications d'alertes et soutenir plus efficacement le travail des analystes SOC Tier-1 dans des environnements opérationnels complexes.

BIBLIOGRAPHIE

Alahmadi, B. A., Axon, L. & Martinovic, I. (2022). 99% False Positives : A Qualitative Study of SOC Analysts' Perspectives on Security Alarms. Dans *Proceedings of the 31st USENIX Security Symposium*, pp. 2783–2800. Repéré à https://www.usenix.org/conference/usenixsecurity22/presentation/alahmadi

Ban, T., Samuel, N., Takahashi, T. & Inoue, D. (2021). Combat Security Alert Fatigue with AI-Assisted Techniques. Dans *2021 IEEE International Conference on Big Data Security on Cloud (BigDataSecurity), High Performance and Smart Computing (HPSC), and Intelligent Data and Security (IDS)*, pp. 9–16. IEEE. doi: 10.1109/BigDataSecurity-HPSC-IDS52210.2021.00010

Bhatt, S., Manadhata, P. K. & Zomlot, L. (2014). The Operational Role of Security Information and Event Management Systems. *IEEE Security&Privacy*, *12*(5), 35–41. doi: 10.1109/MSP.2014.103

Bono, J., Grana, J. & Xu, A. (2024). Generative AI and Security Operations Center Productivity : Evidence from Live Operations. *arXiv preprint arXiv :2411.03116*. Repéré à https://arxiv.org/abs/2411.03116

Freitas, S., Kalajdjieski, J., Gharib, A. & McCann, R. (2025). AI-Driven Guided Response for Security Operation Centers with Microsoft Copilot for Security. Dans *Companion Proceedings of the ACM on Web Conference 2025*, WWW '25, pp. 191–200., New York, NY, USA. Association for Computing Machinery. doi: 10.1145/3701716.3715209

Furnell, S. (2021). The cybersecurity workforce and skills. *Computers&Security*, *100*, 102080. doi: https://doi.org/10.1016/j.cose.2020.102080

Gartner, Inc. (2023, septembre). *Gartner Survey Revealed 34% of Organizations Are Already Using or Implementing AI Application Security Tools*. Press release. Repéré à https://www.gartner.com/en/newsroom/press-releases/2023-09-18-gartner-survey-revealed-34-percent-of-organizations-are-already-using-or-implementing-ai-application-security-tools

Kersten, L., Mulders, T., Zambon, E., Snijders, C. & Allodi, L. (2023). 'Give Me Structure' : Synthesis and Evaluation of a (Network) Threat Analysis Process Supporting

Tier 1 Investigations in a Security Operation Center. Dans *Proceedings of the 19th Symposium on Usable Privacy and Security (SOUPS 2023)*, pp. 97–111. Repéré à https://www.usenix.org/conference/soups2023/presentation/kersten

Motlagh, F. N., Hajizadeh, M., Majd, M., Najafi, P., Cheng, F. & Meinel, C. (2024). Large language models in cybersecurity : State-of-the-art. *arXiv preprint arXiv :2402.00891*.

Mughal, A. A. (2022, 1). Building and Securing the Modern Security Operations Center (SOC). *International Journal of Business Intelligence and Big Data Analytics*, pp. 1–15. Repéré à https://research.tensorgate.org/index.php/IJBIBDA/article/view/21

Muntean, M. & Militaru, F.-D. (2023). Metrics for Evaluating Classification Algorithms. Dans *Education, Research and Business Technologies : Proceedings of the 21st International Conference on Informatics in Economy (IE 2022)*, pp. 307–317. Springer. doi: 10.1007/978-981-19-6755-9_24

Mutemwa, M., Mtsweni, J. & Zimba, L. (2018). Integrating a Security Operations Centre with an Organization's Existing Procedures, Policies and Information Technology Systems. Dans *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, pp. 1–6. doi: 10.1109/ICONIC.2018.8601251

Oniagbi, O. (2024). *Evaluation of LLM Agents for the SOC Tier 1 Analyst Triage Process* [Master's Thesis]. Turku, Finland. Supervisors : Antti Hakkala and Ismayil Hasanov, Repéré à https://www.utupub.fi/handle/10024/178601

Radu, A., Kersten, L., Wosyka, R., Mulders, T., Zambon, E. & Allodi, L. (2025). A Test Tool to Evaluate the Skill Sets of Tier-1 Security Analysts in a SOC Environment : A Case Study from Recruitment to Operations. *Workshop on Security Operation Center Operations and Construction (WOSOC)*. doi: https://dx.doi.org/10.14722/wosoc.2025.23001

Roy, J. & Balde, E. A. (2026). Toward Context-Aware Alert Classification in Security Operations Centers Using LLMs. Dans *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, pp. 1–10. doi: 10.1109/ICAIC67076.2026.11395735

Sharma, K., Kumar, P. & Li, Y. (2025). OG-RAG : Ontology-Grounded Retrieval-Augmented Generation for Large Language Models. Dans *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 32950–32969.

Repéré à https://aclanthology.org/2025.emnlp-main.1674/

Shin, T., Razeghi, Y., Logan IV, R. L., Wallace, E. & Singh, S. (2020). AutoPrompt : Eliciting Knowledge from Language Models with Automatically Generated Prompts. Dans *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4222–4235. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.346

Shyr, C., Ren, B., Hsu, C.-Y., Tinker, R. J., Cassini, T. A., Hamid, R., Wright, A., Bastarache, L., Peterson, J. F. & Malin, B. A. (2025). A Statistical Framework for Evaluating the Repeatability and Reproducibility of Large Language Models. *medRxiv*. Preprint.

Sundaramurthy, S. C., Bardas, A. G., Case, J., Ou, X., Wesch, M., McHugh, J. & Rajagopalan, S. R. (2015). A human capital model for mitigating security analyst burnout. Dans *Eleventh symposium on usable privacy and security (SOUPS 2015)*, pp. 347–359.

Tariq, S., Baruwal Chhetri, M., Nepal, S. & Paris, C. (2025). Alert Fatigue in Security Operations Centres : Research Challenges and Opportunities. *ACM Computing Surveys*, *57*(9), 1–38. doi: 10.1145/3671234

Vielberth, M., Böhm, F., Fichtinger, I. & Pernul, G. (2020). Security Operations Center : A Systematic Study and Open Challenges. *IEEE Access*, *8*, 227756–227779. doi: 10.1109/ACCESS.2020.3045514

Vielberth, M. & Pernul, G. (2018, novembre). A Security Information and Event Management Pattern. Dans *Proceedings of the 12th Latin American Conference on Pattern Languages of Programs (SLPLoP)*. University of Regensburg. doi: 10.5283/epub.41139

Zeadally, S., Adi, E., Baig, Z. & Khan, I. A. (2020). Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity. *IEEE Access*, *8*, 23817–23837. doi: 10.1109/ACCESS.2020.2968045

Zhong, C., Yen, J., Liu, P. & Erbacher, R. F. (2016). Automate Cybersecurity Data Triage by Leveraging Human Analysts' Cognitive Process. Dans *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity)*, pp. 357–363. IEEE. doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.41

APPENDICE A

EXEMPLE COMPLET D'ALERTE SIEM


```
{
  "_index": "wazuh-alerts-4.x-2025.09.11",
  "_id": "YEfdOZkBIDwfn7_zzKqj",
  "_score": null,
  "_source": {
    "input": {
      "type": "log"
    },
    "agent": {
      "ip": "192.168.1.101",
      "name": "DESKTOP-MDNNEJF",
      "id": "001"
    },
    "manager": {
      "name": "ubuntu22"
    },
    "data": {
      "win": {
        "eventdata": {
          "subjectLogonId": "0x3e7",
          "subjectDomainName": "WORKGROUP",
          "targetLinkedLogonId": "0x0",
          "impersonationLevel": "%%1833",
          "authenticationPackageName": "Negotiate",
          "targetLogonId": "0x3e7",
          "logonProcessName": "Advapi",
          "logonGuid": "{00000000-0000-0000-0000-000000000000}",
          "targetUserName": "SYSTEM",
          "keyLength": "0",
          "elevatedToken": "%%1842",
          "subjectUserSid": "S-1-5-18",
          "processId": "0x2dc",
          "processName": "C:\\\\Windows\\\\System32\\\\services.exe",
          "targetDomainName": "NT AUTHORITY",
          "targetUserSid": "S-1-5-18",
          "virtualAccount": "%%1843",
          "logonType": "5",
          "subjectUserName": "DESKTOP-MDNNEJF$"
        },
        "system": {
          "eventID": "4624",
          "keywords": "0x8020000000000000",
          "providerGuid": "{54849625-5478-4994-a5ba-3e3b0328c30d}",
          "level": "0",
          "channel": "Security",
          "opcode": "0",
          "message":"L'ouverture de session d'un compte s'est correctement déroulée.

Objet :
    ID de sécurité:        S-1-5-18
    Nom du compte :        DESKTOP-MDNNEJF$
    Domaine du compte :    WORKGROUP
    ID d'ouverture de session : 0x3E7


Informations d'ouverture de session :
    Type d'ouverture de session :        5
    Mode administrateur restreint :       -
```

```
    Credential Guard à distance :         -
    Compte virtuel :                      Non
    Jeton élevé :                         Oui


Niveau d'emprunt d'identité :
    Emprunt d'identité


Nouvelle ouverture de session :
    ID de sécurité :              S-1-5-18
    Nom du compte :               SYSTEM
    Domaine du compte :            NT AUTHORITY
    ID d'ouverture de session :    0x3E7
    ID d'ouverture de session liée : 0x0
    Nom du compte réseau :         -
    Domaine du compte réseau :     -
    GUID d'ouverture de session :  {00000000-0000-0000-0000-000000000000}


Informations sur le processus :
    ID du processus :      0x2dc
    Nom du processus :     C:\Windows\System32\services.exe


Informations sur le réseau :
    Nom de la station de travail : -
    Adresse du réseau source :     -
    Port source :                  -


Informations détaillées sur l'authentification :
    Processus d'ouverture de session :    Advapi
    Package d'authentification :           Negotiate
    Services en transit :                   -
    Nom du package (NTLM uniquement) :      -
    Longueur de la clé :                    0


Cet événement est généré lors de la création d'une ouverture de session.
Il est généré sur l'ordinateur sur lequel l'ouverture de session a été effectuée.

Le champ Objet indique le compte sur le système local qui a demandé l'ouverture
de session. Il s'agit le plus souvent d'un service, comme le service Serveur,
ou un processus local tel que Winlogon.exe ou Services.exe.

Le champ Type d'ouverture de session indique le type d'ouverture de session
qui s'est produit. Les types les plus courants sont 2 (interactif) et 3 (réseau).

Le champ Nouvelle ouverture de session indique le compte pour lequel la nouvelle
ouverture de session a été créée, par exemple, le compte qui s'est connecté.

Les champs relatifs au réseau indiquent la provenance d'une demande d'ouverture
de session à distance. Le nom de la station de travail n'étant pas toujours
disponible, peut être laissé vide dans certains cas.

Le champ du niveau d'emprunt d'identité indique la portée de l'emprunt
d'identité que peut prendre un processus dans la session d'ouverture de session.

Les champs relatifs aux informations d'authentification fournissent des détails
sur cette demande d'ouverture de session spécifique.
    - Le GUID d'ouverture de session est un identificateur unique pouvant servir
      à associer cet événement à un événement KDC.
    - Les services en transit indiquent les services intermédiaires qui ont
      participé à cette demande d'ouverture de session.
    - Nom du package indique quel est le sous-protocole qui a été utilisé parmi
```

```
      les protocoles NTLM.
    - La longueur de la clé indique la longueur de la clé de session générée.
      Elle a la valeur 0 si aucune clé de session n'a été demandée."

        "version": "3",
        "systemTime": "2025-09-11T17:40:52.7326338Z",
        "eventRecordID": "19804",
        "threadID": "804",
        "computer": "DESKTOP-MDNNEJF",
        "task": "12544",
        "processID": "756",
        "severityValue": "AUDIT_SUCCESS",
        "providerName": "Microsoft-Windows-Security-Auditing"
      }
    }
  },
  "rule": {
    "mail": false,
    "level": 3,
    "pci_dss": [
      "10.2.5"
    ],
    "hipaa": [
      "164.312.b"
    ],
    "tsc": [
      "CC6.8",
      "CC7.2",
      "CC7.3"
    ],
    "description": "Windows Logon Success",
    "groups": [
      "windows",
      "windows_security",
      "authentication_success"
    ],
    "nist_800_53": [
      "AU.14",
      "AC.9"
    ],
    "gdpr": [
      "IV_32.2"
    ],
    "firedtimes": 7,
    "mitre": {
      "technique": [
        "Valid Accounts"
      ],
      "id": [
        "T1078"
      ],
      "tactic": [
        "Defense Evasion",
        "Persistence",
        "Privilege Escalation",
        "Initial Access"
      ]
    },
    "id": "60106",
    "gpg13": [
      "7.1",
      "7.2"
    ]
  },
  "location": "EventChannel",
  "decoder": {
```

```
      "name": "windows_eventchannel"
    },
    "id": "1757612453.1671822",
    "timestamp": "2025-09-11T17:40:53.108+0000"
  },
  "fields": {
    "timestamp": [
      "2025-09-11T17:40:53.108Z"
    ]
  },
  "sort": [
    1757612453108
  ]
}
```

APPENDICE B

CONTEXTE ORGANISATIONNEL COMPLET

```
 {
"meta": {
  "org_name": "ExampleCorp",
  "timezone": "America/Toronto",
  "description": "Contexte enrichi pour triage SOC. Intègre rôles Linux, seuils
  de récurrence et logique de vérification contextuelle."
},

"business_hours": {
  "weekdays": {
    "mon": [["08:30:00", "18:30:00"]],
    "tue": [["08:30:00", "18:30:00"]],
    "wed": [["08:30:00", "18:30:00"]],
    "thu": [["08:30:00", "18:30:00"]],
    "fri": [["08:30:00", "17:30:00"]]
  },
  "holidays": ["2025-12-25", "2025-12-26"]
},

"network": {
  "office_cidrs": ["192.168.1.0/24", "10.10.0.0/16"],
  "trusted_remote_vpn": ["203.0.113.0/28"],
  "external_allowlist": ["8.8.8.8/32"]
},

"groups": {
  "IT":        { "members": ["it_admin", "helpdesk01", "ops_lead", "env-admin", "root"],
  "can_remote": true,  "allowed_out_of_hours": true,  "allowed_from_external": true,
  "notes": "Opérations/Support"},
  "ADMINS":    { "members": ["sec_lead", "cto"] ,
  "can_remote": true,  "allowed_out_of_hours": true,  "allowed_from_external": true,
  "notes": "Administrateurs sécurité/exec" },
  "FINANCE":   { "members": ["acct_1", "acct_manager", "ARTUser", "art"],
  "can_remote": false, "allowed_out_of_hours": false, "allowed_from_external": false,
  "notes": "Comptabilité"  },
  "HR":        { "members": ["hr_1", "hr_manager"],
  "can_remote": false, "allowed_out_of_hours": false, "allowed_from_external": false,
  "notes": "Ressources humaines"  },
  "SERVICE_ACCOUNTS": { "members": ["SYSTEM", "LOCAL SERVICE", "NETWORK SERVICE",
  "svc_backup"] , "is_technical": true, "can_remote": false, "allowed_out_of_hours": true,
  "allowed_from_external": false, "notes": "Identités techniques/non humaines" }
},

"users": {
  "it_admin":    { "groups": ["IT", "ADMINS"], "roles": ["sysadmin"] },
  "helpdesk01":  { "groups": ["IT"], "roles": ["support"] },
  "ops_lead":    { "groups": ["IT"], "roles": ["operations"] },
  "env-admin":   { "groups": ["IT"], "roles": ["linux_admin","windows_admin"] },
  "root":        { "groups": ["IT"], "roles": ["root_admin"] },
  "sec_lead":    { "groups": ["ADMINS"], "roles": ["security_admin"] },
  "cto":         { "groups": ["ADMINS"], "roles": ["exec"] },
  "acct_1":      { "groups": ["FINANCE"], "roles": ["accountant"] },
  "acct_manager":{ "groups": ["FINANCE"], "roles": ["finance_manager"] },
  "ARTUser":     { "groups": ["FINANCE"], "roles": ["finance_analyst"] },
  "art":         { "groups": ["FINANCE"], "roles": ["finance_support"] },
  "hr_1":        { "groups": ["HR"], "roles": ["hr_associate"] },
  "hr_manager":  { "groups": ["HR"], "roles": ["hr_manager"] },
  "SYSTEM":      { "groups": ["SERVICE_ACCOUNTS"], "roles": ["system"] },
```

```
    "LOCAL SERVICE":   { "groups": ["SERVICE_ACCOUNTS"], "roles": ["system"] },
    "NETWORK SERVICE": { "groups": ["SERVICE_ACCOUNTS"], "roles": ["system"] },
    "svc_backup":      { "groups": ["SERVICE_ACCOUNTS"], "roles": ["backup_service"] }
  },

  "assets": {
    "localhost.localdomain": { "role": "linux_workstation", "site": "HQ",
    "owner_group": "IT", "tags": ["linux", "pam"] },
    "DESKTOP-MDNNEJF":       { "role": "windows_workstation", "site": "HQ",
    "owner_group": "IT", "tags": ["windows", "workgroup"] },
    "SRV-DB-01":             { "role": "server", "site": "HQ",
    "owner_group": "FINANCE", "tags": ["windows", "db"] }
  },

  "role_catalog": {
    "linux_workstation": {
      "expected_technical_accounts": ["root", "env-admin"],
      "expected_service_processes": ["/usr/sbin/cron", "/usr/sbin/chpasswd"],
      "typical_logon_types": [0, 1]
    },
    "windows_workstation": {
      "expected_technical_accounts": ["SYSTEM", "LOCAL SERVICE", "NETWORK SERVICE"],
      "expected_service_processes": ["C:\\\\Windows\\\\System32\\\\services.exe",
      "C:\\\\Windows\\\\System32\\\\svchost.exe"],
      "typical_logon_types": [2, 3, 5]
    },
    "server": {
      "expected_technical_accounts": ["SYSTEM", "svc_backup", "root"],
      "expected_service_processes": ["C:\\\\Windows\\\\System32\\\\services.exe",
      "/usr/sbin/sshd"],
      "typical_logon_types": [3, 5]
    }
  },

  "geo": {
    "allowed_countries": ["CA", "US"]
  },

  "maintenance_windows": [
    { "start": "2025-10-20T22:00:00-04:00", "end": "2025-10-21T02:00:00-04:00",
    "note": "Nightly patching" }
  ],

  "policies": {
    "recurrent_events": {
      "description": "Politique générique de détection d'événements répétés
      (auth, accès, process, etc.).",
      "enabled": true,

      "threshold": 4,
      "window_minutes": 5,

      "group_by": ["user", "host", "event_type"],
      "ignore_accounts": ["SYSTEM", "LOCAL SERVICE", "NETWORK SERVICE", "svc_backup"],
      "count_service_accounts": false,
      "reset_on_success": true,

      "default_single_event_policy": "not-interesting",

      "decision_logic": [
        "Étape 1: Évaluer le contexte (utilisateur connu/droit, plage horaire, site/IP,
        pays, rôle de l'actif).",
        "Étape 2: Si une règle critique est violée (utilisateur inconnu, IP hors office/
        vpn/allowlist, pays interdit), classer 'interesting' immédiatement.",
        "Étape 3: Sinon, un seul événement isolé reste 'not-interesting' par défaut.",
        "Étape 4: Si le même event_type se répète >= threshold dans window_minutes pour
```

```
        la même clé (user+host+event_type), classer 'interesting'."
      ]
    },

    "event_type_hints": {
      "description": "Aides génériques pour dériver event_type sans dépendre
      d'une plateforme.",
      "rules": [
        { "match": {"rule.groups": "authentication_failed"},
        "event_type": "authentication_failed" },
        { "match": {"win.system.eventID": [4625]},
        "event_type": "authentication_failed" },
        { "match": {"rule.description": "login failed"},
        "event_type": "authentication_failed" },
        { "match": {"rule.description": "authentication failure"},
        "event_type": "authentication_failed" }
      ]
    }
  },

"policies": {
  "auth_failures": {
    "description": "Politique de triage des échecs d'authentification avec seuils de
    récurrence et exceptions contextuelles.",
    "recurrence": { "threshold": 4, "window": "5min" }
    },
    "notes": "Un seul échec d'authentification = not-interesting. Devient interesting
    si >= threshold dans window OU si exception."
},

  "facts": {
    "password_change": {
      "sensitive": true,
      "typically_initiated_by": ["user_self_service", "it_admin"],
      "requires_verification_for_groups": ["FINANCE", "HR"]
    },
    "sudo_actions": {
      "critical_roles": ["root", "env-admin"],
      "review_required_for": ["FINANCE", "HR", "GUEST"]
    }
  }
}
```

APPENDICE C

JEU DE DONNÉES COMPLET ET VÉRITÉ TERRAIN DES ALERTES

TABLEAU C.1 : Vérité terrain complète des alertes Windows

Alert ID	Description	ID	Technique	Tactic	Ground Truth
1757612453.1671822	Windows Logon Success	T1078	Valid Accounts	Defense Evasion, Persistence, Privilege Escalation, Initial Access	Not Interesting
1757612118.1653432	Powershell process spawned Windows command shell instance	T1059.003	Windows Command Shell	Execution	Interesting
1757613591.1821187	Service startup type changed	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757613583.1810657	Software protection service scheduled successfully	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757612118.1647649	Suspicious Windows cmd shell execution	T1087, T1059.003	Account Discovery, Windows Command Shell	Discovery, Execution	Not Interesting
1757612081.1635793	Powershell process created an executable file in Windows root folder	T1105	Ingress Tool Transfer	Command and Control	Interesting
1757612080.1632975	Executable file dropped in folder commonly used by malware	T1105	Ingress Tool Transfer	Command and Control	Interesting
1757612033.1626234	Windows command prompt started by an abnormal process	T1059.003	Windows Command Shell	Execution	Interesting
1757611970.1617034	PowerShell process creation with suspicious command line	T1070.004	File Deletion	Defense Evasion	Interesting
1757611774.1589886	Registry modification by env-admin	T1027, T1112	Obfuscated Files or Information, Modify Registry	Defense Evasion	Interesting
1757611590.1558176	PowerShell created a new scripting file under User data folder	T1105, T1059	Ingress Tool Transfer, Command and Scripting Interpreter	Command and Control, Execution	Interesting
1757611525.1549468	Suspicious Process - lsm.exe	T1055	Process Injection	Defense Evasion, Privilege Escalation	Interesting
1757613944.1876416	Registry Value Added to System	T1112	Modify Registry	Defense Evasion	Interesting
1757611467.1517244	Command shell started script with /c modifier	T1059	Command and Scripting Interpreter	Execution	Interesting
1757611465.1409874	Executable dropped in Windows root folder	T1570	Lateral Tool Transfer	Lateral Movement	Interesting
1757611399.1392111	Windows backup and restore tool sdclt.exe launched via PowerShell with High integrity level	T1548, T1059.003	Abuse Elevation Control Mechanism, Windows Command Shell	Privilege Escalation, Defense Evasion, Execution	Interesting
1757611185.1319274	Powershell process invoked known auto-elevated utility ComputerDefaults.EXE	T1548.002	Bypass User Account Control	Privilege Escalation, Defense Evasion	Interesting
1757611185.1335881	Potential UAC bypass via registry modification	T1548.002, T1112	Bypass User Account Control, Modify Registry	Privilege Escalation, Defense Evasion	Interesting
1757611047.1222736	Powershell executed a base64 encoded command	T1059.001	PowerShell	Execution	Interesting
1757610816.1173186	Powershell script compiling code using CSC.exe	T1027.004	Compile After Delivery	Defense Evasion	Interesting
1757610654.1112794	Suspicious Process - lsass	T1055	Process Injection	Defense Evasion, Privilege Escalation	Interesting
1757610608.1101539	Executable file dropped in Users/Public folder	T1105	Ingress Tool Transfer	Command and Control	Interesting
1757607634.840437	Powershell process created executable file in AppData temp folder	T1105	Ingress Tool Transfer	Command and Control	Interesting
1757605861.812840	License activation failed on DESKTOP-MDNNEJF	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757613920.1840104	Registry Key Integrity Checksum Changed	T1565.001, T1112	Stored Data Manipulation, Modify Registry	Impact, Defense Evasion	Not Interesting
1757657178.471270	Registry Value Entry Deleted	T1070.004, T1485, T1112	File Deletion, Data Destruction, Modify Registry	Defense Evasion, Impact	Interesting
1757656769.388902	Net command execution	T1087	Account Discovery	Discovery	Not Interesting
1757946519.877972	Non-service account logged off	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757946519.865952	Windows Workstation Logon Success	T1078	Valid Accounts	Defense Evasion, Persistence, Privilege Escalation, Initial Access	Not Interesting
1757946508.845453	Logon Failure - Unknown user or bad password	T1531	Account Access Removal	Impact	Not Interesting
1757946519.862756	Special privileges assigned to new logon	T1484	Domain Policy Modification	Defense Evasion, Privilege Escalation	Interesting
1757945940.790942	Software protection service scheduled successfully	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757915969.377664	Discovery activity executed	T1087	Account Discovery	Discovery	Not Interesting
1757797573.1708146	VSS service shutdown due to idle timeout	SYSTEM	SYSTEM	SYSTEM	Not Interesting

TABLEAU C.2 : Vérité terrain complète des alertes Linux

Alert ID	Description	ID	Tecchnique	Tactic	Ground Truth
1757448450.1607664	PAM : User changed password	SYSTEM	SYSTEM	SYSTEM	Interesting
1757421591.908968	PAM : Login session closed for user root	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757421591.908504	PAM Login session opened for user root by env-admin	T1078	Valid Accounts	Defense Evasion, Persistence, Privilege Escalation, Initial Access	Not Interesting
1761280350.215525	User authentication failure	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1761280233.214804	Wazuh agent started	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757439593.1276633	Integrity checksum changed for /etc/subuid	T1565.001	Stored Data Manipulation	Impact	Interesting
1757439233.1227872	nftables table existence check passed	T1562, T1562.004	Disable or Modify System Firewall, Impair Defenses	Command and Control	Not Interesting
1757396591.314337	New group added to the system	SYSTEM	SYSTEM	SYSTEM	Interesting
1757390403.187211	Apparmor DENIED	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757439593.1274699	File added to /etc/cron.d/persistevil	SYSTEM	SYSTEM	SYSTEM	Interesting
1757302739.245548	Systemd service exited due to failure	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757448450.1606656	Successful sudo to ROOT executed	T1548.003	Sudo and Sudo Caching	Privilege Escalation, Defense Evasion	Not Interesting
1757448450.1605639	New user added to the system	T1136	Create Account	Persistence	Interesting
1757439237.1234145	SCA summary : CIS Ubuntu Linux 22.04 LTS Benchmark v2.0.0. : Score less than 50% (44)	SYSTEM	SYSTEM	SYSTEM	Interesting
1759427649.2846606	PAM : User login failed	T1110.001	Password Guessing	Credential Access	Not Interesting
1757072746.14293	Vulnerability CVE-2023-28755 affecting ruby-rubygems was solved	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757069230.9758	New dpkg package installed	SYSTEM	SYSTEM	SYSTEM	Not Interesting
1757556911.71617	Failed attempt to run sudo	T1548.003	Sudo and Sudo Caching	Privilege Escalation, Defense Evasion	Interesting
1756921222.1803596	Failed scan for permissions on /etc/shells	T1003, T1003.008, T1222, T1222.002	OS Credential Dumping, /etc/passwd and /etc/shadow, File and Directory Permissions Modification, Linux and Mac File and Directory Permissions Modification	Defense Evasion	Interesting
1756921222.1806864	Failed scan for permissions on /etc/security/opasswd	T1003, T1003.008, T1222, T1222.002	OS Credential Dumping, /etc/passwd and /etc/shadow, File and Directory Permissions Modification, Linux and Mac File and Directory Permissions Modification	Defense Evasion	Not Interesting
1758964101.591531	File deletion	T1070.004, T1485	File Deletion, Data Destruction	Defense Evasion, Impact	Not Interesting