

الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbas



Engineering Thesis

To obtain the diploma of **Engineering Degree**

Field of Study: **Computer Science**

Specialization: **Intelligence Artificielle et Science de Données (IASD)**

Theme

AURA-MAS: Agentic and Multi-Agent Intelligent Surveillance

A Privacy-Aware, Edge-First Multi-Agent System for Multimodal Event Detection,

Inter-Sensor Coordination, and Explainable Alerting

Presented by
ZEDDOUN Lokmane

Defended on: **23 September, 2026**

In front of the jury composed of:

Pr. RAHMOUN Abdellatif
Dr. BENDELLA Mohammed Salih
Dr. BENSENANE Hamdane

President
Examiner
Supervisor

Academic Year: 2025/2026

Abstract

Conventional video surveillance often sends many camera streams to a control room, where a small team must notice rare critical events. This thesis presents **AURA-MAS** (Agentic Understanding & Response Architecture), a privacy-aware, edge-first *hierarchical multi-agent system* for site surveillance. The system detects security-relevant events, combines related evidence, can request verification, and presents the result to a human operator.

At the edge, video agents use YOLO11 detection, ByteTrack tracking, zone rules, and optional CLIP-based anomaly scoring. Audio agents classify sound events and fall back to signal processing when required. They publish structured events rather than raw footage on an MQTT/Redis-compatible bus. A fusion agent groups events into hypotheses with a noisy-OR confidence score. For uncertain hypotheses, the coordinator can run a Contract Net-inspired auction to select a camera for verification. A deterministic policy agent applies thresholds, severity rules, and cooldowns. An explanation layer runs after the alert decision and checks every cited evidence identifier against the event log before presenting a report.

The Python implementation was evaluated in a 373-run campaign over nine annotated replay scenarios. Paired Wilcoxon tests with Holm–Bonferroni correction found no statistically significant event-level F1 difference among the four architectures. The auction is implemented, but its field-of-view term was inactive during measurement, so the campaign does not validate the intended spatial allocation. Cross-modal fusion occurs, but repeated evidence from one camera saturates noisy-OR before its confidence effect can be isolated. The explanation guardrail rejects unsupported evidence references, and the privacy pipeline anonymizes exported evidence while keeping an append-only decision log. The thesis therefore places most confidence in its evaluation apparatus: isolated subprocesses, repeated runs, a preserved pre-fix baseline, and a written methodology-change log.

Keywords— Intelligent Surveillance, Multi-Agent Systems, Agentic AI, Large Language Models, Object Detection, Multimodal Fusion, Auction-Based Coordination, Contract Net Protocol, Edge Computing, Explainable AI, Privacy by Design, Video Anomaly Detection

Résumé

La vidéosurveillance conventionnelle demeure largement centralisée et passive : les flux de nombreuses caméras convergent vers une salle de contrôle où des opérateurs surchargés doivent remarquer des événements critiques rares, un paradigme qui passe mal à l'échelle et réagit lentement. Cette thèse propose et évalue **AURA-MAS**, un *système multi-agents hiérarchique*, respectueux de la vie privée et orienté *edge*, pour la surveillance intelligente de sites : il détecte, corrobore, vérifie et explique les événements de sécurité tout en maintenant l'opérateur humain au centre de la décision.

Au niveau de la périphérie, des agents de perception autonomes traitent la vidéo (détection YOLO11 avec suivi multi-objets ByteTrack, règles de zones déclaratives et score d'anomalie sémantique zero-shot fondé sur CLIP) et l'audio (classification d'événements sonores avec repli en traitement du signal), en publiant des événements structurés — jamais de flux bruts — sur un bus de messages léger MQTT/Redis. Au niveau de la coordination, un agent de fusion agrège les événements en hypothèses d'incident par fusion tardive spatio-temporelle avec combinaison de confiance noisy-OR ; un agent coordinateur résout les hypothèses incertaines par allocation de tâches aux enchères inspirée du protocole Contract Net, en chargeant la caméra la mieux placée de vérifier activement l'incident ; et un agent de politique déterministe applique des seuils auditables, une cartographie de sévérité et une hystérésis d'alerte. Une couche d'explication *agentique* encadrée par des règles, fondée sur les grands modèles de langage, transforme ensuite les alertes acceptées en rapports d'incident en langage naturel dont les citations de preuves sont vérifiées mécaniquement par rapport au journal d'événements réel.

Le système complet est implémenté en Python et évalué par une campagne de 373 exécutions sur neuf scénarios de jeu annotés. Des tests de Wilcoxon appariés avec correction de Holm-Bonferroni ne mettent en évidence aucune différence statistiquement significative entre les quatre architectures pour la F1 au niveau événement. Le mécanisme d'enchères est implémenté, mais son terme informatif de recouvrement de champ de vision était inactif durant la mesure : il n'est donc pas validé tel qu'il est spécifié. La fusion multimodale se déclenche, mais son effet pratique sur la confiance ne peut être isolé car des preuves répétées de la même caméra saturent le noisy-OR. Le garde-fou d'explication et le pipeline de protection de la vie privée constituent des contributions structurelles : le premier rejette les références de preuves non étayées, tandis que le second maintient l'anonymisation des preuves exportées et une journalisation des décisions en ajout seul. La contribution empirique la plus solide est ainsi l'appareil d'évaluation lui-même : isolation par sous-processus, répétitions, conservation d'une base pré-correction et journal écrit des changements méthodologiques.

Mots-clés : Surveillance intelligente, Systèmes multi-agents, IA agentique, Grands modèles de langage, Détection d'objets, Fusion multimodale, Coordination par enchères, Protocole Contract Net, Informatique en périphérie, IA explicable, Protection de la vie privée dès la conception, Détection d'anomalies vidéo

ملخص

لا تزال المراقبة بالفيديو التقليدية مركزية وسلبية إلى حد كبير: تتدفق بثوث كاميرات عديدة نحو غرفة تحكم يتعين فيها على مشغلين مثقلين ملاحظة أحداث حرجة نادرة، وهو نموذج ضعيف التوسع وبطيء الاستجابة. تقترح هذه الأطروحة وتقيم نظام AURA-MAS، وهو نظام متعدد الوكلاء هرمي يحترم الخصوصية ويعمل على حافة الشبكة، للمراقبة الذكية للمواقع: يكتشف الأحداث الأمنية ويعززها ويتحقق منها ويشرحها مع إبقاء المشغل البشري في قلب القرار.

على مستوى الحافة، يعالج وكلاء إدراك مستقلون الفيديو كشف YOLO11 مع تتبع متعدد الأجسام ByteTrack وقواعد مناطق تصريحية ودرجة شذوذ دلالية بلا أمثلة قائمة على CLIP والصوت وتصنيف الأحداث الصوتية مع بديل قائم على معالجة الإشارة، وينشرون أحداثاً مهيكلية --- وليس تدفقات خام أبداً --- عبر ناقل رسائل خفيف MQTT/Redis. وعلى مستوى التنسيق، يجمع وكيل الدمج الأحداث في فرضيات حوادث عبر دمج متأخر زمني-مكاني، ويحل وكيل التنسيق الفرضيات غير المؤكدة عبر تخصيص المهام بالمزاد المستوحى من بروتوكول Contract Net، بتكليف الكاميرا الأنسب بالتحقق النشط من الحادث، بينما يطبق وكيل السياسة الحتمي عتبات قابلة للتدقيق وخريطة خطورة وتباطؤ الإنذارات. ثم تحوّل طبقة شرح وكيالية مفيدة بالقواعد ومبنية على النماذج اللغوية الكبيرة الإنذارات المقبولة إلى تقارير حوادث بلغة طبيعية تُفحص استشهاداتها بالأدلة آلياً مقابل سجل الأحداث الفعلي.

نُفذ النظام الكامل بلغة بايثون وقيم من خلال حملة تضم 373 تشغيلاً عبر تسعة سيناريوهات إعادة تشغيل مشروحة. ولا تُظهر اختبارات Wilcoxon المزدوجة مع تصحيح Holm-Bonferroni فروقاً ذات دلالة إحصائية بين البنى الأربع في مقياس F1 على مستوى الحدث. آلية المزاد منفذة، لكن حذّها المعلوماتي لتداخل مجال الرؤية كان غير نشط أثناء القياس، ومن ثم لم تُثبت صلاحيتها كما حُدّت. يعمل الدمج متعدد الوسائط، لكن لا يمكن عزل أثره العملي على الثقة لأن الأدلة المتكررة من الكاميرا نفسها تُشبع نموذج noisy-OR. وتمثل حماية شرح التنبيهات وخط أنابيب الخصوصية مساهمتين بنيويتين: الأول يرفض مراجع الأدلة غير المدعومة، والثاني يحافظ على إخفاء هوية الأدلة المصدّرة وسجل قرارات تراكمي. لذلك فإن أقوى مساهمة تجريبية هي جهاز التقييم نفسه: عزل العمليات الفرعية، والتكرارات، وخط أساس محفوظ قبل الإصلاح، وسجل مكتوب لتغيرات المنهجية.

الكلمات المفتاحية: المراقبة الذكية، الأنظمة متعددة الوكلاء، الذكاء الاصطناعي الوكيلي، النماذج اللغوية الكبيرة، كشف الأجسام، الدمج متعدد الوسائط، التنسيق بالمزاد، بروتوكول Contract Net، الحوسبة الطرفية، الذكاء الاصطناعي القابل للتفسير، الخصوصية منذ التصميم، كشف الشذوذ في الفيديو.