

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 ٨ ماي ١٩٤٥ - سيدي بلعباس
Higher School of Computer Science 8 Mai 1945 - Sidi Bel Abbès



Thesis

To obtain the diploma of **Engineering Degree**

Field of Study: **Computer Science**

Specialization: **Artificial Intelligence and data science**

Theme

Agentic AutoMLOps for Hospitals: A Self-Managing Platform for Medical AI Model Monitoring and Deployment

Presented by
Yahiaoui Yamina

Defended on: **[24 September, 2026]**

In front of the jury composed of

Dr. DIF Nassima
Dr. Malki Abdelhamid
Dr. Malki Mimoun
Dr. IRATNI Naima

President of the Jury
Thesis Supervisor
Co-Supervisor
Examiner

Academic Year: 2025/2026

Abstract

Hospitals are deploying deep learning models to support clinical work, but a model that is accurate on the day it is released does not stay accurate. Imaging devices are replaced, acquisition protocols change, and patient populations shift, so production data slowly moves away from the training data and performance degrades silently. Keeping even a small fleet of models healthy therefore requires a permanent team of specialists, which most hospitals do not have. Recent agentic systems, in which several language-model agents cooperate to carry out the work of a data-science team, suggest that part of this burden could be automated.

This thesis examines that possibility. It first assembles the conceptual foundation the question requires, bringing together literatures that are usually treated separately: deep learning for medical imaging, the machine learning lifecycle, AutoML, MLOps, model degradation and data drift, large language models, and agentic AI. It then reviews six agentic systems published in 2025 and 2026, each analysed on the same grid of architecture, evaluation protocol, and results, and each assessed for the strength of the evidence it provides.

The review yields a clear picture. These systems converge on a common shape — specialised agents around a language-model core, numerical optimisation delegated to deterministic tools, generated code fenced in by templates and repair loops, and memory that turns out to be load-bearing — and they disagree mainly on scope and on the place given to the human. Their common weakness is the same: agentic automation of model construction is advancing quickly, while the operation of deployed models is nearly untouched. Only one of the six treats drift, on synthetic data; two include no human oversight; and none ties an operational decision to a metric chosen for the clinical cost of its errors.

From this gap we derive the requirements of an agentic AutoMLOps platform for hospitals: monitoring that combines immediate label-independent drift signals with delayed label-dependent performance metrics, a closed set of permitted actions, deterministic safety rules anchored on the false-negative rate, and human approval of every action that affects production — a design in which the language model advises and never decides.

Clinova is the platform built to those requirements. It supervises a fleet of medical-imaging models already in service through three layers: a pipeline that prepares the data, searches the hyperparameters, and admits a new version only if its false-negative rate does not regress against the model currently in production; a controller that runs a governed observe–think–act loop over each model, detects drift on live traffic without waiting for labels, and draws every proposal from a closed set of six actions that deterministic rules confirm or override; and a console that gives an administrator, an engineer and a clinician three different views of the same system, and lets a clinician dispute a prediction without that dispute moving a model on its own.

Keywords: MLOps, AutoML, agentic AI, large language models, data drift, model monitoring, medical imaging, human-in-the-loop, continuous training.

ملخص

تعتمد المستشفيات بشكل متزايد على نماذج التعلم العميق لمساعدة الأطباء في تحليل الصور الطبية، غير أن النموذج الدقيق يوم إطلاقه لا يبقى كذلك. فأجهزة التصوير تستبدل، والبروتوكولات تتغير، وخصائص المرضى تتبدل، مما يجعل بيانات الإنتاج تبتعد تدريجياً عن بيانات التدريب فيتراجع الأداء دون أن يلاحظ. لذلك يتطلب الحفاظ على عدد قليل من النماذج في الخدمة فريقاً دائماً من المختصين، وهو ما لا تملكه أغلب المستشفيات. وقد أظهرت الأنظمة الوكيلية الحديثة، التي تتعاون فيها عدة وكلاء قائمين على نماذج لغوية لأداء عمل فريق من علماء البيانات، أن جزءاً من هذا العبء قابل للأتمتة.

تدرس هذه المذكرة هذه الإمكانيات. فهي تبني أولاً الأساس المفاهيمي اللازم، بجمع مجالات تعالج عادة بشكل منفصل: التعلم العميق في التصوير الطبي، ودورة حياة التعلم الآلي، والتعلم الآلي المؤتمت، وعمليات تعلم الآلة، وتدهور النماذج وانحراف البيانات، والنماذج اللغوية الكبيرة، والذكاء الاصطناعي الوكيل. ثم تستعرض ستة أنظمة وكيلية نشرت بين سنتي ٢٠٢٠ و٢٠٢٢، حللت وفق الشبكة نفسها من حيث البنية وبروتوكول التقييم والنتائج، مع تقدير مدى متانة الأدلة التي يقدمها كل منها.

وتظهر هذه المراجعة صورة واضحة: تتقارب هذه الأنظمة في بنيتها العامة، وتختلف أساساً في مجال تدخلها وفي المكانة الممنوحة للإنسان، إذ تتقدم أتمتة بناء النماذج بسرعة بينما يبقى تشغيل النماذج المنشورة شبه مهملة. فنظام واحد فقط من الستة يعالج انحراف البيانات، وعلى بيانات اصطناعية؛ ونظامان لا يتضمنان أي إشراف بشري؛ ولا يربط أي منها قراراً تشغيلياً بمقياس مختار وفق الكلفة السريرية لأخطائه.

انطلاقاً من هذه الضجوة، نستخلص متطلبات منصة أوتوملوبس وكيلية موجهة للمستشفيات: مراقبة تجمع بين إشارات انحراف فورية لا تحتاج إلى تسميات ومقاييس أداء متأخرة، ومجموعة مغلقة من الإجراءات المسموح بها، وقواعد أمان صارمة تستند إلى معدل السلبيات الكاذبة، ومصادقة بشرية على كل إجراء يمس الإنتاج، بحيث يقدم النموذج اللغوي المشورة ولا يتخذ القرار.

«كلينوفا» هي المنصة المبنية وفق هذه المتطلبات. تشرف على مجموعة من نماذج التصوير الطبي في الخدمة عبر ثلاث طبقات: مسار يهيئ البيانات، ويبحث آلياً في المعاملات الفائقة، ولا يقبل نسخة جديدة إلا إذا لم يتراجع فيها معدل السلبيات الكاذبة مقارنةً بالنموذج العامل في الإنتاج؛ ومتحكم ينفذ على كل نموذج حلقة محكومة من الملاحظة والتفكير والفعل، يكشف انحراف البيانات من حركة الطلبات الحية دون انتظار التسميات، ويستمد كل اقتراح من مجموعة مغلقة من ستة إجراءات تؤكد قواعدها أمان صارمة أو تنقضها؛ وواجهة تمنح المسؤول والمهندس والطبيب ثلاث رؤى مختلفة للنظام نفسه، وتتيح للطبيب الاعتراض على تنبؤ دون أن يحرك اعتراضه أي نموذج بمفرده.

الكلمات المفتاحية: عمليات تعلم الآلة، التعلم الآلي المؤتمت، الذكاء الاصطناعي الوكيل، النماذج اللغوية الكبيرة، انحراف البيانات، مراقبة النماذج، التصوير الطبي، الإنسان في الحلقة، التدريب المستمر.