# Artificial Intelligence in French Healthcare and Emergency Response: A Multi-Site Case Study from Predictive Modeling to Generative AI

Roxane Elias Mallouhy[1] and Christophe Guyeux[2]

[1]Al Yamamah University, Saudi Arabia
r_mallouhy@yu.edu.sa.
[2]Université Marie et Louis Pasteur, CNRS, institut FEMTO-ST, F-90000 Belfort, France
christophe.guyeux@univ-fcomte.fr.

This paper presents two operational case studies of Artificial Intelligence (AI) deployment in French healthcare and emergency response systems: the ARS Bourgogne-Franche-Comté (BFC) predictive analytics platform and the Marcus/IAppel emergency call management project. The ARS BFC initiative, operational since 2018 across approximately fifty healthcare facilities, employs XGBoost-based forecasting models achieving Mean Absolute Error (MAE) of 5–6 admissions/day for hospitalizations and 12–13 admissions/day for overall ED visits across 1 to 14-day horizons. The platform integrates digital twin technology for simulation-based strategic planning. The Marcus project, piloted in the Ain department with 40 real emergency calls analyzed, implements a multi-stage AI pipeline combining automatic speech recognition (Whisper), named entity recognition, and large language model classification, achieving F1-scores exceeding 93% for urgency detection. Both initiatives demonstrate a human-augmentation paradigm where AI assists rather than replaces operators. We document the technical architectures, quantitative performance metrics, user feedback, and lessons learned from multi-year deployments, addressing practical challenges including data governance, algorithmic limitations for rare events, and regulatory alignment with the EU AI Act.

**Keywords:** Artificial Intelligence, Healthcare Systems, Emergency Response, Predictive Modeling, Digital Twin, Large Language Models, Retrieval-Augmented Generation, Civil Security

# 1 Introduction

The rapid advancement of Artificial Intelligence (AI) over the past decade has fundamentally transformed numerous sectors, with healthcare emerging as one of the most promising domains for impactful applications. From diagnostic support systems to predictive analytics and natural language processing (NLP) for clinical documentation, AI technologies are increasingly permeating medical practice and health system management. This evolution is driven by the convergence of three key factors: the exponential growth in computational capacity, the availability of large-scale health datasets, and significant methodological breakthroughs in machine learning and deep learning architectures.

Within the European context, France has positioned itself as a leading actor in the deployment of AI for healthcare and emergency response systems [4, 5]. Following the strategic recommendations of the Villani Report [11], substantial investments have been directed toward research infrastructure, data governance frameworks, and collaborative initiatives between academic institutions, hospitals, and regional health agencies (Agences Régionales de Santé, ARS). This national commitment reflects a broader recognition that AI can address critical challenges facing modern healthcare systems: aging populations, resource constraints, and the imperative for more personalized and preventive care models [12].

The spectrum of AI applications in French healthcare is remarkably diverse. Machine learning and deep learning algorithms are now employed for predictive modeling of hospital admissions and disease progression [7]. NLP techniques enable the automated extraction of clinically relevant information from unstructured electronic health records [8]. Computer vision systems enhance the accuracy and speed of medical image interpretation, supporting radiologists in detecting anomalies across X-rays, CT scans, and MRIs [9]. Collectively, these technologies are enabling a transition from reactive to anticipatory healthcare delivery.

This paper presents a comprehensive examination of AI deployment in the French healthcare ecosystem through two flagship regional initiatives. The first, led by ARS Bourgogne-Franche-Comté, encompasses predictive models for emergency department admissions, digital twin technology for simulation-based strategic planning, and operational decision support systems. The second, the Marcus project conducted in the Ain department, focuses on generative AI and retrieval-augmented generation (RAG) for intelligent emergency call triage and doctrinal support in civil security contexts. Beyond documenting these implementations, this paper critically examines the challenges inherent to responsible AI deployment, including data governance, algorithmic bias, and the preservation of human oversight in high-stakes medical decisions.

The remainder of this paper is organized as follows. Section 2 positions our work relative to prior research in healthcare forecasting, emergency call AI, and digital twins. Section 3 presents Case Study 1: the ARS Bourgogne-Franche-Comté predictive analytics initiative, including forecasting models, the PredictOps platform, and digital twin technology. Section 4 presents Case Study 2: the Marcus/IAppel project for AI-assisted emergency call management. Section 5 synthesizes lessons learned across both deployments. Finally, Section 6 concludes with perspectives for future development.

## 2 Related Work

The deployment of AI in healthcare operations has generated substantial research across three domains directly relevant to our case studies.

***Forecasting Hospital Demand.***

Time series forecasting for emergency department admissions has evolved from traditional statistical approaches (ARIMA, exponential smoothing) toward machine learning methods. Recent work demonstrates that gradient boosting (XGBoost, LightGBM) and deep learning architectures (LSTM, Temporal Fusion Transformers) can outperform classical methods, particularly when incorporating exogenous variables such as meteorological conditions, epidemiological indicators, and calendar effects [7]. However, performance degrades substantially for rare events and activity peaks, a challenge documented across multiple healthcare forecasting studies.

***AI in Emergency Call Centers.***

Several commercial and research initiatives have explored AI integration in emergency dispatch. Carbyne and Corti provide AI-enhanced call enrichment and triage support, while the NG112/NG911 infrastructure enables IP-based multimedia emergency communications. Academic work has demonstrated the feasibility of speech-to-text transcription, speaker diarization, and named entity recognition for emergency calls [1, 2]. Privacy-preserving approaches, including differential privacy and de-identification protocols, address the sensitive nature of emergency communications [10]. The EU AI Act classifies emergency call triage systems as high-risk, imposing requirements for human oversight, transparency, and robustness.

***Digital Twins for Healthcare Planning.***

Digital twin technology—creating virtual replicas of physical systems for simulation and optimization—has emerged as a strategic planning tool for healthcare infrastructure [6]. Applications include capacity planning, resource allocation optimization, and scenario analysis for emergency preparedness. Unlike static coverage analyses, digital twins enable dynamic simulation that adapts to evolving resource configurations and demand patterns.

## 3 Case Study 1: AI-Powered Forecasting for Hospital Resource Optimization

This section presents the first case study: an AI-driven initiative developed in collaboration with ARS Bourgogne-Franche-Comté (ARS BFC), the regional health agency responsible for coordinating healthcare services across eastern France. The Bourgogne-Franche-Comté region spans approximately 48,000 square kilometers with a population of 2.8 million, encompassing both major university hospitals and rural healthcare facilities.

***Project Timeline and Governance.***

The initiative originated from an operational challenge identified at SDIS 25 (Doubs Fire and Rescue Service): the structural “scissors effect” of increasing operational demand against stagnating resources. While the project began within the fire-rescue domain, its methodology and forecasting capabilities proved directly applicable to hospital capacity planning—leading to a strategic partnership with ARS Bourgogne-Franche-Comté (the regional health agency) to extend the approach across the healthcare continuum. Key milestones include:

- **November 2018:** Project inception at SDIS 25 and feasibility study
- **January 2019:** Validation by CASDIS (departmental fire service board)
- **2019–2020:** Scientific research phase, partnership with ARS BFC established
- **2020:** Proof-of-concept development for hospital forecasting
- **2021:** Pre-industrialization and beta deployment at CODIS 25
- **2023–present:** Multi-site expansion across ARS BFC territory (healthcare facilities)

The project has produced seven peer-reviewed publications and three additional manuscripts under review, with media coverage in La Gazette des Communes (June 2025). This cross-sectoral governance model—originating in emergency services and extending to healthcare—illustrates how forecasting methodologies can transfer between domains sharing common operational challenges.

The initiative now comprises three interconnected components: (i) XGBoost-based forecasting models predicting emergency department admissions and bed requirements across approximately fifty healthcare facilities, (ii) the PredictOps platform providing real-time decision support for hospital administrators, and (iii) digital twin technology for simulation-based strategic planning.

## 3.1 Forecasting Emergency Department Admissions and Bed Demand

The forecasting component of this initiative addresses a fundamental challenge in hospital operations: the mismatch between unpredictable patient demand and resource capacity that must be planned in advance. Emergency departments face particularly acute manifestations of this challenge, as admission patterns exhibit both systematic variation (weekday/weekend cycles, seasonal flu epidemics, holiday periods) and stochastic fluctuations that complicate resource allocation decisions.

The predictive modeling effort leverages the RPU dataset (Résumés de Passage aux Urgences—Emergency Department Visit Summaries), a standardized data collection mandated across French emergency departments. This dataset provides granular information on patient arrivals, acuity levels, and disposition outcomes, enabling the development of models that anticipate the Minimum Daily Bed Requirement (Besoin Journalier Minimal en Lits). The overarching objective is to facilitate a transition from reactive crisis management—where resources are mobilized in response to emerging overcrowding—to proactive capacity planning that anticipates demand fluctuations before they materialize.

The forecasting framework addresses two complementary prediction targets:

- **Short-term ED admissions:** 3-day moving average predictions enabling immediate operational adjustments
- **Medium-term hospitalization volume:** Weekly average predictions supporting staffing and scheduling decisions

These predictions support several concrete operational use cases. Human resources departments can anticipate staffing requirements and adjust personnel scheduling ahead of predicted demand surges, particularly during school vacation periods when both patient volumes and staff availability fluctuate. Elective admission scheduling can be optimized to avoid planned hospitalizations during forecasted high-activity periods. Discharge planning teams can prioritize Friday discharges when weekend demand is predicted to be elevated, maximizing bed availability during peak periods.

To validate model performance, multiple deep learning architectures were evaluated across four forecasting horizons: 3, 7, 15, and 30 days. The comparative analysis focused on Dijon University Hospital (CHU Dijon), the region's largest tertiary care center, using data spanning the 2023 calendar year. Figures 1–4 present the comparison between predicted and observed ED admission volumes, with Root Mean Squared Error (RMSE) metrics quantifying prediction accuracy for each model configuration and time horizon.

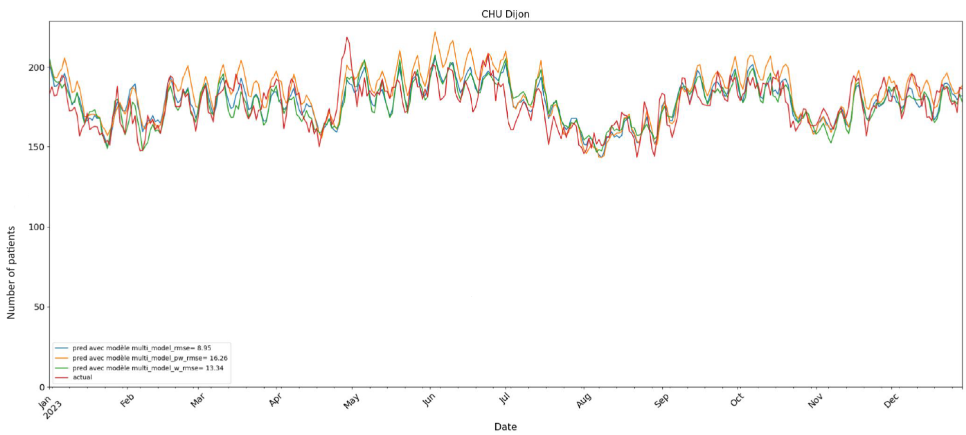


**Fig. 1** 3-day horizon forecast for daily ED admissions, CHU Dijon, 2023. Blue: observed values; Orange: XGBoost prediction. RMSE = 3.97 admissions/day.

***Technical Implementation and Results.***

The forecasting engine employs XGBoost (eXtreme Gradient Boosting). Key hyperparameters include reduced `min_child_weight` to improve sensitivity to extreme values, conservative `learning_rate` for stability, and tuned `subsample`/`colsample_bytree` for regularization. For **daily-level predictions** (1 to 14-day horizons), Table 1

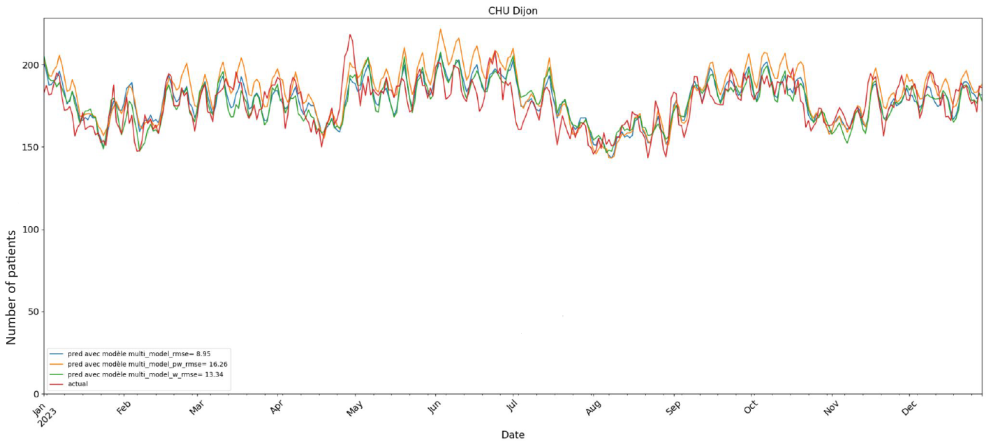


**Fig. 2** 7-day horizon forecast for daily ED admissions, CHU Dijon, 2023. Blue: observed values; Orange: XGBoost prediction. RMSE = 4.32 admissions/day.

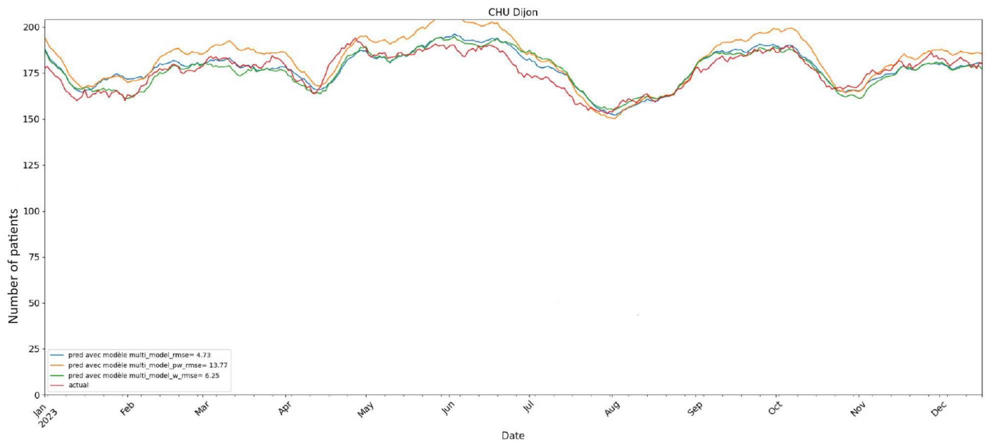


**Fig. 3** 15-day horizon forecast for daily ED admissions, CHU Dijon, 2023. Blue: observed values; Orange: XGBoost prediction. RMSE = 8.44 admissions/day.

presents Mean Absolute Error (MAE) across prediction horizons and patient categories for the ARS BFC deployment.

A complementary proof-of-concept study [3] validated the XGBoost approach at **hourly granularity** on 10 years of retrospective data (2010–2019) from two French emergency departments (Hôpital Mercy and Hôpital Bel Air), demonstrating consistent superiority over traditional baselines. Table 2 compares model performance for hourly admission prediction—note that hourly MAE values are naturally lower than daily MAE due to the finer temporal resolution.

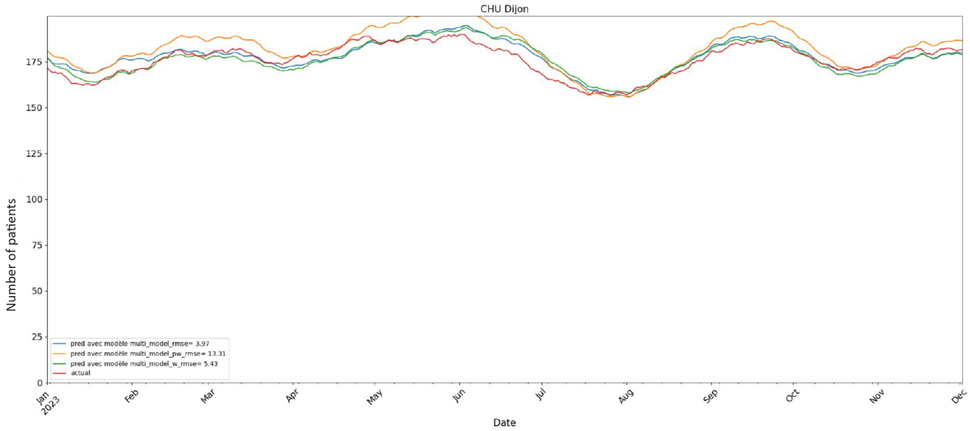


**Fig. 4** 30-day horizon forecast for daily ED admissions, CHU Dijon, 2023. Blue: observed values; Orange: XGBoost prediction. RMSE = 16.26 admissions/day.

**Table 1** Daily Prediction MAE by Horizon and Patient Category (CHU Dijon, 2023). Values represent mean absolute error in admissions per day.

| Horizon | ED Admissions (All) | Hospitalizations | Senior Hospitalizations |
|---|---|---|---|
| 1 day | 12.32 | 5.21 | 5.8 |
| 3 days | 12.56 | 5.34 | 6.0 |
| 7 days | 12.75 | 5.27 | 5.9 |
| 14 days | – | 5.41 | 6.2 |

**Table 2** Hourly Prediction Model Comparison (MAE in patients per hour, Brossard et al. 2025)

| Model | Mercy | Bel Air |
|---|---|---|
| Seasonal Average | ∼6.5 | ∼6.0 |
| ARIMA | ∼4.0 | ∼3.0 |
| LASSO | 2.90 | 2.90 |
| LightGBM | 2.66 | 2.66 |
| **XGBoost (optimized)** | **2.63** | **2.64** |

The multi-site study employed 712 explanatory variables per hourly prediction, including 672 lagged historical values (4 weeks × 7 days × 24 hours), calendar features, meteorological data, traffic patterns, and epidemiological indicators. Beyond approximately 125 variables, performance plateaued then degraded—confirming the importance of regularization over feature accumulation. The operational MAE of ±2.6 patients/hour represents a significant improvement over traditional approaches and is sufficient for practical staffing decisions.

Performance degrades substantially for rare events and activity peaks. With limited historical data for extreme events, statistical power remains insufficient. Aggressive hyperparameter tuning to capture peaks led to unacceptable false positive rates, degrading user confidence—a documented limitation that informs ongoing research priorities.

## 3.2 Operational Decision Support: The PredictOps Platform

Building upon the forecasting models described above, the initiative has developed an integrated decision support platform—PredictOps—designed to translate predictive insights into actionable operational intelligence. This platform addresses a critical gap between algorithmic predictions and practical decision-making: while forecasting models can anticipate demand patterns, effective resource management requires contextualizing these predictions within the complex operational environment of healthcare delivery.

#### *Multi-Source Data Integration.*

The PredictOps architecture integrates two complementary categories of predictive variables. Endogenous variables derive from internal operational data streams: intervention logs from emergency services, incoming call volumes at dispatch centers, real-time resource availability, duty rosters, and specialist schedules. These internally-generated data provide granular visibility into current operational capacity and constraints. Exogenous variables capture external environmental and societal factors influencing healthcare demand: meteorological conditions (temperature extremes, precipitation), public holidays and school schedules, ambient pollution levels, traffic patterns, epidemiological surveillance data, and environmental risk indicators including drought conditions and fire risk assessments. The integration of these heterogeneous data sources enables predictions that account for both internal operational dynamics and external demand drivers.

#### *Operational Use Cases.*

The enhanced predictive capabilities support several concrete applications extending beyond traditional bed management. Call center dimensioning can be optimized by allocating staff based on forecasted call volumes, urgency distributions, and historical patterns of demand variation. Non-emergency patient transport services can be dynamically scheduled to avoid conflicts with predicted periods of high emergency demand. Emergency vehicle pre-positioning strategies can be informed by forward-looking risk projections rather than solely reactive dispatch protocols. Patient routing decisions—particularly referrals to secondary care centers—can incorporate real-time predictions of facility workload, resource availability, and geographic accessibility constraints.

#### *Complementary Planning Horizons.*

The platform architecture distinguishes between two temporal decision-making layers. PredictOps addresses the immediate operational horizon, providing real-time decision support for demand anticipation over hours to days. OptimOps addresses the strategic planning horizon, analyzing long-term patterns to inform policy development, infrastructure investment, and structural resource allocation. This dual-layer architecture enables healthcare organizations to operate effectively in both reactive and anticipatory modes, responding to immediate operational pressures while simultaneously pursuing longer-term optimization objectives.

***Performance Monitoring Framework.***

Effective operational decision support requires continuous monitoring of performance indicators that reflect both efficiency and equity dimensions:

- **Response Time Metrics:** Elapsed time from emergency call to first responder arrival, assessed across different pre-positioning configurations and compared against regulatory thresholds.
- **Service Availability:** Duration of center unavailability due to resource constraints (personnel, vehicles, equipment), directly impacting response capacity.
- **Capacity Utilization:** Frequency with which operational units approach or exceed capacity limits, identifying systematic bottlenecks requiring structural intervention.
- **Geographic Coverage Equity:** Continuous monitoring of underserved areas to ensure equitable access to emergency services across the regional territory.

***Spatial Accessibility Analysis.***

To assess geographic equity in emergency service access, the project conducted detailed spatial analysis of municipal coverage. Cartographic visualizations (Figures 5–8) illustrate cumulative catchment areas as emergency centers are progressively added to the service network. The resulting maps employ a gradient scale—with red indicating underserved municipalities and blue representing areas with optimal coverage—enabling identification of persistent gaps in geographic accessibility and informing strategic decisions regarding infrastructure expansion or resource redeployment.

## 3.3 Strategic Planning Through Digital Twin Technology

The third component of the ARS BFC initiative extends the analytical framework from operational decision support to strategic planning through digital twin technology [6]. A digital twin constitutes a high-fidelity virtual replica of a real-world system, continuously synchronized with operational data to mirror system behavior, resource states, and interaction dynamics. In the context of emergency services, digital twins enable stakeholders to simulate alternative configurations, test policy interventions, and evaluate strategic decisions without disrupting actual service delivery.

The digital twin developed for the ARS BFC project provides a comprehensive virtual representation of the regional emergency response network. Unlike static coverage analyses derived from historical data or geographic information systems, this dynamic simulation environment continuously adapts to evolving resource configurations, personnel availability, and population distribution patterns. This capability transforms strategic planning from a periodic, document-driven exercise into an interactive, data-informed process.

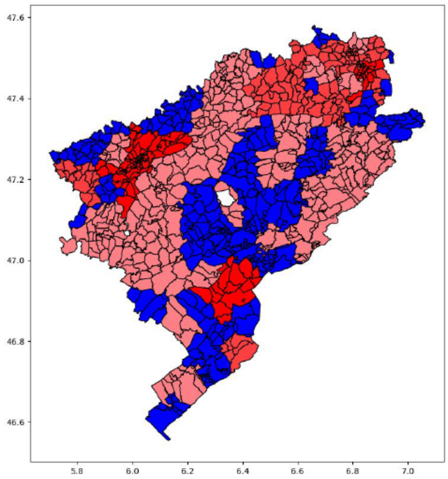


**Fig. 5** Center 1

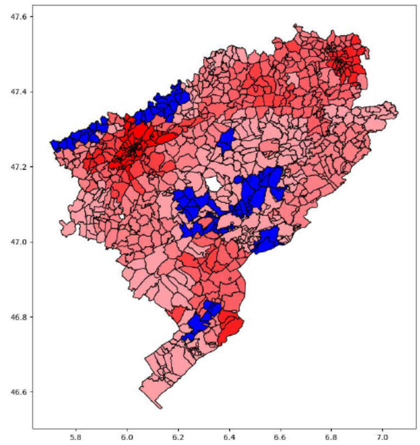


**Fig. 6** Centers 1+2

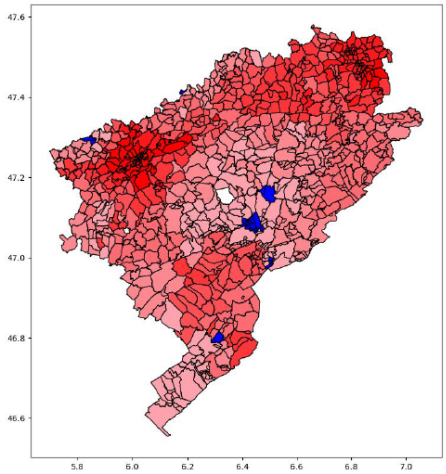


**Fig. 7** Centers 1+2+3

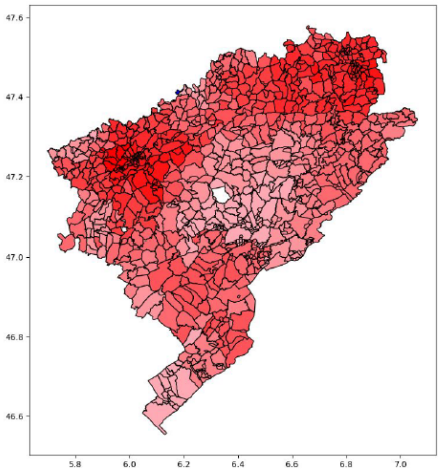


**Fig. 8** Centers 1 to 4

***Integration with Regulatory Planning Frameworks.***

The digital twin supports alignment with the SDACR (Schéma Départemental d'Analyse et de Couverture des Risques—Departmental Plan for Risk Analysis and Coverage), the mandatory regulatory framework governing emergency service deployment in French departments. Simulation capabilities enable planners to evaluate proposed modifications to service delivery structures against regulatory coverage requirements before implementation, identifying potential compliance gaps or optimization opportunities.

***Operational Configuration Testing.***

The simulation environment—operating in silico—enables systematic evaluation of operational modifications without real-world disruption. Representative applications include: reconfiguring emergency vehicle equipment allocations to optimize match between available resources and anticipated mission profiles; designing targeted training programs to address identified skill gaps within specific geographic areas or risk categories; and identifying volunteer recruitment priorities based on projected demand patterns and demographic characteristics of underserved zones.

***Counterfactual Analysis of Historical Events.***

The platform enables retrospective simulation of past emergency events under alternative resource configurations. By recreating historical scenarios with modified personnel deployments, equipment allocations, or dispatch protocols, decision-makers can develop evidence-based insights into how alternative approaches might have affected response times, patient outcomes, or resource utilization. This capability supports organizational learning and continuous improvement processes.

***Optimization Under Resource Constraints.***

Whether operating under steady-state conditions or responding to evolving resource constraints, the digital twin enables systematic exploration of trade-offs between competing objectives. Multi-objective optimization algorithms can identify Pareto-efficient configurations that balance response time minimization, coverage equity, cost containment, and personnel workload distribution.

***Seasonal and Event-Driven Planning.***

The simulation platform accommodates temporal variation in demand patterns, enabling proactive resource reallocation in response to predictable seasonal factors—tourist influx during vacation periods, weather-related risk elevations, or scheduled large-scale events. This anticipatory planning capability ensures that resource positioning reflects projected demand rather than historical averages.

***Quantitative Impact Assessment.***

Preliminary simulations demonstrated tangible operational benefits. For instance, simulating the temporary closure of one emergency center predicted a 12% increase in average response times for the affected catchment area—a projection subsequently validated against actual data from a planned maintenance period. Such counterfactual analyses enable evidence-based capacity planning decisions.

The digital twin thus complements the operational systems described previously, extending the analytical horizon from immediate resource management to long-term strategic governance. Collectively, the forecasting models, PredictOps platform, and digital twin technology represent an integrated approach to intelligent healthcare resource management—one that demonstrates how AI can augment not only daily clinical operations but also system-wide planning and policy development across the healthcare continuum.

# 4 Case Study 2: Generative AI for Emergency Call Management

This section presents the Marcus/IAppel project, a pilot initiative conducted in the French department of Ain exploring generative AI integration into emergency call management. The project represents a collaboration between SDIS 01 (Service Départemental d'Incendie et de Secours de l'Ain), DTNum (Direction de la Transformation Numérique), and the Université de Franche-Comté.

***Project Timeline and Team.***

The initiative evolved from academic research (2019–2022) to operational experimentation (2025):

- **2019–2022:** Foundational research on privacy-preserving audio classification (PhD thesis by M. Abi Kanaan)
- **2024:** System integration and EENA 112 Conference presentation (Canary Islands, December 2024)
- **February 2025:** Demonstration milestone at SDIS 01
- **March–June 2025:** Field experimentation with 40 real emergency calls analyzed
- **April 2025:** EENA Conference presentation (Helsinki)

The project team includes: CDT Guillaume Royer-Fey (product owner, SDIS 01), Enzo Bellicaud (full-stack developer), David Laiymani (FEMTO-ST), with institutional support from DTNum (Nathalie Guicherd, Soumia Dermouche) and the Ministry of Interior's digital transformation directorate.

***Data and Performance.***

Training data comprises 2,315 emergency calls (4 days of recordings), with approximately 158 calls retained after quality filtering (non-first-responder calls excluded). Mean call duration after preprocessing is approximately 4 minutes. The system achieves:

- **Urgency detection:** F1-score $> 93\%$
- **Category classification:** F1-score $\approx 74\%$ across 12 emergency categories
- **Target latency:** $< 500$ms end-to-end for critical entity extraction

## 4.1 Technical Architecture

The Marcus architecture implements a multi-stage processing pipeline [1]:

1. **Speech-to-Text:** Whisper (OpenAI) for real-time transcription, with VOSK as fallback for robustness
2. **Speaker Diarization:** Pyannote.Audio for caller/operator separation (DER $\approx$ 24% on French ETAPE corpus)
3. **Named Entity Recognition:** Extraction of names, addresses, phone numbers, and temporal references
4. **LLM Classification:** Urgency prediction (binary) and category classification (12 classes: fire, traffic accident, medical emergency, etc.)

5. **Decision Tree Execution:** Explainable routing through domain-specific decision trees
6. **Summary Generation:** Operator-oriented call synopsis for rapid situational awareness

Noise filtering at the initial stage identifies and routes non-emergency communications: accidental pocket dials, prank calls, and inquiries better directed to non-emergency services. AI-based classification distinguishes legitimate emergency calls, reducing operator workload.

For legitimate calls, Named Entity Recognition algorithms extract structured data from unstructured speech [10]. Speaker diarization maintains dialogue coherence for subsequent analysis. These components transform raw audio into structured, actionable information augmenting operator situational awareness.

## 4.2 Operator Feedback and Adoption

Semi-structured interviews were conducted with N=8 SDIS 01 operators during the March–June 2025 experimentation phase. Qualitative feedback highlights key adoption factors:

> "This experimentation makes the application of AI in our professions more concrete." (Operator feedback, 2025)

> "The initial results on transcription and summaries are encouraging." (Operator feedback, 2025)

> "It is essential that the tool remains at the service of the operator." (Operator feedback, 2025)

These testimonials underscore a consistent theme: AI is perceived as an assistant augmenting human capabilities, not as a replacement for professional judgment. The explainability of decision trees—which display the reasoning path for each classification—was cited as a critical factor for user trust.

## 4.3 Structured Decision Support for Triage

Following AI preprocessing, calls are routed through a standardized decision framework aligned with predefined emergency categories: medical emergencies, traffic accidents, fires, public safety incidents, and environmental hazards. Structured decision trees guide operators through systematic question sequences, ensuring consistent triage across operators and accelerating the determination of appropriate response levels and intervention types.

The final routing decision—dispatching emergency medical services, fire-rescue units, or law enforcement—remains subject to human validation, but operators benefit from AI-generated contextual summaries that synthesize caller information, extracted entities, and preliminary classification assessments. This hybrid human-AI architecture maintains the reliability and accountability of human judgment while leveraging AI's capacity for rapid information processing and pattern recognition.

Specialized decision trees have been developed for natural hazard events, including floods, storms, and landslides [2]. These frameworks incorporate contextual factors—presence of endangered persons, fire propagation risk, involvement of animals—to generate urgency classifications (critical, urgent, routine) and recommend appropriate response resources. The systematic structure ensures that calls are routed to appropriate services with minimal delay while maintaining decision consistency across the heterogeneous call types characteristic of unified emergency operations.

## 4.4 Retrieval-Augmented Generation for Doctrinal Support

Beyond call processing, the Marcus Project deploys Retrieval-Augmented Generation (RAG) technology to provide contextual and doctrinal support for civil security operations. RAG architecture combines the generative capabilities of LLMs with targeted retrieval from validated document repositories, enabling responses that integrate both model knowledge and authoritative content from regulatory, procedural, and training materials.

The RAG prototype developed for civil security officers supports four primary use cases:

***National Doctrine Guidance.***

Operators can pose complex situational queries—for instance, requesting recommended actions for an acetylene cylinder fire in a residential building—and receive responses synthesized from national emergency procedures and technical guidance documents.

***Local Doctrine Adaptation.***

The system assists regional authorities in adapting national-level frameworks to local operational contexts, accounting for geographic specificities, resource availability, and jurisdictional arrangements.

***Training and Professional Development.***

Officers in training interact with conversational agents built from institutional materials, including thesis reports from ENSOSP (École Nationale Supérieure des Officiers de Sapeurs-Pompiers), enabling personalized, on-demand learning experiences that supplement formal training programs.

***Technical Support Automation.***

A specialized conversational agent addresses technical support inquiries, providing automated responses where feasible and assisting in ticket preparation for issues requiring human intervention.

These applications demonstrate how RAG technology, when coupled with generative AI, enables secure, context-aware systems capable of supporting doctrine interpretation, knowledge transfer, and operational decision-making. Whether enhancing situational awareness during active emergencies, accelerating professional development, or streamlining administrative processes, the Marcus Project illustrates how

LLMs can be responsibly operationalized in high-stakes public safety contexts while maintaining appropriate human oversight.

## 5 Lessons Learned and Practical Recommendations

Beyond technical contributions, both case studies yield actionable insights for organizations deploying AI in healthcare and emergency response contexts.

***Stakeholder Engagement as Success Factor.***

Both initiatives demonstrate that successful AI deployment requires sustained, multi-stakeholder collaboration. The Marcus project's co-construction model—involving SDIS 01 operators, DTNum engineers, FEMTO-ST researchers, and Ministry representatives—proved essential for ensuring both technical validity and operational acceptability. Similarly, the ARS BFC initiative benefited from early engagement with hospital administrators, clinical staff, and regional health authorities.

***Explainability as Adoption Driver.***

Operator feedback consistently emphasized the importance of AI transparency. In Marcus, the decision tree architecture explicitly traces the reasoning path for each classification, enabling operators to understand and validate AI recommendations. This explainability was cited as a critical factor differentiating acceptable AI assistance from opaque automation.

***Managing Expectations for Rare Events.***

The ARS BFC project revealed fundamental limitations in predicting activity peaks. With limited historical data (few years, single institution), statistical power for extreme events remains insufficient. Aggressive XGBoost hyperparameter tuning to capture peaks led to unacceptable false positive rates, degrading user confidence. This finding informs realistic expectation-setting for healthcare forecasting systems.

***Privacy and Regulatory Alignment.***

Both projects anticipate EU AI Act requirements for high-risk systems (emergency call triage is explicitly classified as high-risk). Key compliance measures implemented include: documented risk management frameworks, data governance protocols ensuring GDPR compliance, comprehensive logging for auditability, human oversight mechanisms preserving operator authority, and robustness testing against adversarial inputs and edge cases.

***From Pilot to Scale.***

Transitioning from successful pilot implementations to broader deployment requires attention to: infrastructure heterogeneity across sites, data quality variation, local workflow adaptations, and change management processes. Both projects document explicit scaling strategies informed by lessons from initial deployments.

## 6 Conclusion and Future Perspectives

This paper has presented a comprehensive examination of artificial intelligence deployment in French healthcare and emergency response systems, drawing on two flagship regional initiatives that demonstrate the transition of AI from research prototype to operational reality.

***Summary of Contributions.***

The ARS Bourgogne-Franche-Comté initiative illustrates how gradient boosting-based forecasting, operationalized through the PredictOps platform, can transform hospital resource management from reactive crisis response to anticipatory capacity planning. Prediction models achieving MAE values between 5 and 13 admissions/day across 1 to 14-day horizons provide actionable intelligence for staffing decisions, elective admission scheduling, and discharge planning. The digital twin component extends this analytical framework to strategic planning, enabling simulation-based evaluation of emergency service configurations without disrupting operational delivery.

The Marcus Project demonstrates the potential of generative AI and retrieval-augmented generation in emergency call management. By integrating natural language processing for call transcription and entity extraction, structured decision trees for triage standardization, and RAG-based systems for doctrinal support, the project establishes a template for AI-augmented emergency operations that preserves human oversight while enhancing operator capabilities.

***Cross-Cutting Insights.***

Across both initiatives, several themes emerge. First, effective AI deployment requires integration with existing operational workflows rather than replacement of established processes. The hybrid human-AI architectures implemented in both projects maintain human decision authority for high-stakes choices while leveraging AI for information processing, pattern recognition, and decision support. Second, prediction and simulation capabilities are most valuable when embedded in platforms designed for operational use—the transition from algorithmic output to actionable intelligence requires careful attention to interface design, workflow integration, and user training.

***Persistent Challenges.***

These advances, while demonstrating substantial potential, do not resolve fundamental challenges inherent to AI deployment in healthcare contexts. Data governance frameworks must evolve to accommodate the data-intensive requirements of AI systems while protecting patient privacy and maintaining public trust. Algorithmic bias remains a concern, particularly given documented instances of differential performance across demographic groups. Digital infrastructure disparities risk concentrating AI benefits in well-resourced urban centers, potentially exacerbating existing health inequities. The preservation of therapeutic relationships and clinical expertise in increasingly automated environments requires ongoing attention.

***Future Directions.***
Several priorities emerge for continued development. Expansion beyond pilot regions will require attention to generalizability across diverse healthcare contexts and population characteristics. Enhanced interoperability between hospital information systems, emergency dispatch platforms, and predictive analytics infrastructure remains a technical priority. Research on hybrid human-AI collaboration models should inform the design of systems that augment rather than supplant professional expertise. Explainable AI approaches will be essential for maintaining transparency and accountability as these systems become more deeply embedded in clinical and operational decision chains.

Ultimately, artificial intelligence in healthcare and civil security represents not a replacement for professional judgment but an amplification of human capabilities. When developed with appropriate attention to ethical considerations, deployed with robust governance frameworks, and integrated thoughtfully into operational workflows, these technologies hold substantial promise for more responsive, equitable, and effective public health infrastructure.

# Acknowledgments

This work is supported by Al Yamamah University. The authors extend their sincere thanks to the College of Engineering for its continuous assistance and institutional support. This research is also partially funded by the ANR Labcom SADIAND project, whose contribution has been instrumental in advancing the study's objectives.

# References

[1] Marianne Abi Kanaan, Jean-François Couchot, Christophe Guyeux, David Laiymani, Talar Atechian, and Rony Darazi. A methodology for emergency calls severity prediction: from pre-processing to bert-based classifiers. In *IFIP international conference on artificial intelligence applications and innovations*, pages 329–342. Springer, 2023.

[2] Marianne Abi Kanaan, Jean-François Couchot, Christophe Guyeux, David Laiymani, Talar Atechian, and Rony Darazi. Combining a multi-feature neural network with multi-task learning for emergency calls severity prediction. *Array*, 21:100333, 2024.

[3] Cyrielle Brossard, Christophe Goetz, Pierre Catoire, Lauriane Cipolat, Christophe Guyeux, Cédric Gil Jardine, Mahuna Akplogan, and Laure Abensur Vuillaume. Predicting emergency department admissions using a machine-learning algorithm: a proof of concept with retrospective study. *BMC Emergency Medicine*, 25(1):1–12, 2025.

[4] Davide Callegarin and Patrick Callier. Enjeux du déploiement de l'intelligence artificielle en santé. *Actualités Pharmaceutiques*, 60(611):21–24, 2021.

[5] Dominique Desbois. French-style applications of artificial intelligence to human health in a european context. In *IFIP International Conference on Human Choice and Computers*, pages 110–122. Springer, 2024.

[6] Hyman Duan, Su Gao, Xu Yang, and Yuanlin Li. The development of a digital twin concept system. *Digital Twin*, 2(10):10, 2025.

[7] Alexander Meyer, Dina Zverinski, Boris Pfahringer, Jörg Kempfert, Titus Kuehne, Simon H Sündermann, Christof Stamm, Thomas Hofmann, Volkmar Falk, and Carsten Eickhoff. Machine learning for real-time prediction of complications in critical care: a retrospective study. *The Lancet Respiratory Medicine*, 6(12):905–914, 2018.

[8] R Nagalakshmi, Surbhi Bhatia Khan, TR Mahesh, Mohammad Alojail, Saurabh Raj Sangwan, Mo Saraee, et al. Enhancing drug discovery and patient care through advanced analytics with the power of nlp and machine learning in pharmaceutical data interpretation. *SLAS technology*, 31:100238, 2025.

[9] Ravindra Sharma, Narendra Kumar, and Vinod Sharma. Computer vision for disease detection—an overview of how computer vision techniques can be used to detect diseases in medical images, such as x-rays and mris. *AI in Disease Detection: Advancements and Applications*, pages 77–98, 2025.

[10] Yakini Tchouka, Jean-François Couchot, David Laiymani, Philippe Selles, and Azzedine Rahmani. Differentially private de-identifying textual medical document is compliant with challenging nlp analyses: Example of privacy-preserving icd-10 code association. *Intelligent Systems with Applications*, 23:200416, 2024.

[11] Cédric Villani, Yann Bonnet, Bertrand Rondepierre, et al. *For a meaningful artificial intelligence: Towards a French and European strategy.* Conseil national du numérique, 2018.

[12] Jenna Wiens and Erica S Shenoy. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, 66(1): 149–153, 2018.