

PhD Research Proposal

Safe Multimodal AI for Cardiology Decision Support and Constrained Agentic Assistance

Applicant: Amdjed Mohammed Said KHELIFI

Email: khelifi.amdjed@univ-oeb.dz **Phone:** +213 549 59 91 82

1. 1. Research Background and Motivation

Cardiology decisions depend on integrating imaging, structured clinical variables, and clinical text. In real workflows these inputs are incomplete, heterogeneous across hospitals, and must be interpreted under uncertainty. This creates an opportunity for AI systems that fuse evidence, quantify uncertainty, and support clinicians without overstepping clinical responsibility.

My long-term goal is to develop clinically deployable AI that improves risk stratification and decision support in cardiology, while meeting requirements for interpretability, calibration, safety, and accountability. I hold a BSc in Computer Science (University of Constantine 2, 2024) and I am completing an MSc in Artificial Intelligence and Data Science (University of Oum El Bouaghi, Algeria). I have built end-to-end machine learning systems across computer vision and NLP, including medical imaging prototypes and multimodal clinical prediction models.

I recently completed a multimodal cardiology research project entitled *A Multimodal Deep Learning Approach for Cardiac Risk Assessment: Integrating Cardiac MRI and Clinical Data*. This work has been accepted for presentation at the SSD International Multi-Conference. It combines a CNN-based imaging branch and a structured-data model, includes interpretability tools, and proposes a clinician-facing prototype. The project clarified what is required for a PhD-level contribution: robustness under missing modalities and domain shift, calibrated uncertainty for risk decisions, and rigorous clinical evaluation beyond internal accuracy.

2. 2. Problem Statement and Research Gaps

Despite progress in medical AI, three gaps remain critical for safe deployment of multimodal and agentic systems:

- 1. Robust multimodal fusion under missingness and domain shift.** Medical imaging protocols, scanners, and populations vary across institutions. Clinical data is often missing, delayed, or noisy. Many multimodal models assume complete inputs and do not degrade gracefully.
- 2. Uncertainty estimation and calibration for clinical risk.** Probability outputs are frequently miscalibrated, which undermines threshold-based triage and can lead to unsafe overconfidence. Clinical deployment requires calibrated risk scores and risk-aware decision rules.

3. **Safe, constrained agentic assistance with accountability.** Agentic AI can draft summaries or propose next steps, but clinical use requires restricted action spaces, auditable reasoning traces, mandatory clinician confirmation, and monitoring for out-of-scope outputs.

3. 3. Aim, Objectives, and Research Questions

3.1. 3.1 Aim

To develop robust multimodal cardiology decision support that remains reliable under missing modalities and real-world shift, and to augment it with a constrained agentic module that assists clinicians in defined workflows with auditable safety controls.

3.2. 3.2 Objectives

1. Design fusion architectures that tolerate incomplete modalities and remain interpretable.
2. Integrate uncertainty estimation and calibration to produce clinically meaningful risk scores.
3. Build a constrained agentic assistant for limited tasks (structured summarization, follow-up suggestion, triage support) with human-in-the-loop control and complete audit logs.
4. Validate clinical utility through technical evaluation and workflow-centred human studies.

3.3. 3.3 Research Questions

1. Which fusion strategies (late fusion, middle fusion, quality-gated attention) provide the best robustness under missing modalities and domain shift?
2. How can uncertainty estimation and calibration improve safety for high-risk minority cases in cardiology prediction?
3. What constraints and audit mechanisms are sufficient to make an agent helpful while remaining controllable and clinically appropriate?
4. Do multimodal predictions combined with constrained assistance reduce clinician workload or errors in realistic simulation studies?

4. 4. Methodology

4.1. 4.1 Data, Modalities, and Target Task

The primary use case is cardiology risk assessment using:

- **Imaging:** cardiac MRI (including cine sequences when available).
- **Structured data:** demographics, vitals, laboratory values, comorbidities, and clinically relevant indicators.
- **Clinical text (optional):** short reports or notes for summarization and rationale generation, subject to approvals and de-identification.

4.2. 4.2 Robust Multimodal Modeling and Fusion

I will begin with a reproducible baseline consistent with my prior work: a deep imaging encoder (CNN or sequence model) and a structured-data model (gradient boosting and/or neural tabular encoder), followed by fusion and a meta-classifier. PhD-level extensions will include:

- **Quality-aware fusion:** learned gating that weights modalities based on quality and missingness indicators.
- **Missing-modality training:** modality dropout and conditional fusion to ensure graceful degradation.
- **Temporal modeling:** when cine MRI is available, evaluate 3D CNN or transformer-based sequence encoders to capture cardiac dynamics.
- **Shift-aware evaluation:** structured stress tests that simulate scanner/protocol shift and realistic missingness patterns.

4.3. 4.3 Uncertainty, Calibration, and Risk-Aware Decision Rules

I will integrate uncertainty estimation (deep ensembles or Bayesian approximations) and calibration (temperature scaling, isotonic regression, and/or conformal prediction when appropriate). Outputs will include calibrated risk scores and confidence tags, enabling safer thresholds and escalation rules (e.g., “high risk with low confidence” triggers additional review rather than automated recommendations).

4.4. 4.4 Interpretability, Monitoring, and Auditability

Interpretability will be integrated for each modality: imaging saliency/activation maps and structured feature attributions. I will study explanation stability under shift and implement versioned logging of predictions, uncertainty, explanations, and model metadata to support monitoring, audits, and error analysis.

4.5. 4.5 Constrained Agentic Assistance with Reinforcement Learning

I will build an agentic module that operates above the predictor. The agent will produce structured summaries and follow-up suggestions using protocol-aligned templates and a restricted action space. Where feasible, it will be trained or tuned using offline

reinforcement learning signals derived from clinician-approved trajectories; otherwise, expert-designed reward shaping within a simulator will be used. Safety constraints will include:

- mandatory clinician confirmation before any recommendation is applied,
- rule-based checks and forbidden-action lists,
- prompt/output logging with traceable rationale,
- monitoring for hallucination and out-of-scope requests.

5. 5. Evaluation and Validation

Technical evaluation will include AUROC, sensitivity/specificity, F1 for imbalance, calibration error, and robustness under missingness and shift. Human-centred evaluation will use simulated clinical tasks measuring time-to-decision, omission errors, override rates, and perceived usefulness. When collaborations and approvals permit, I will prioritize external validation on an independent cohort and report generalization transparently.

6. 6. Work Plan and Timeline (4 Years)

Year 1: literature review; data agreements; baseline reproduction; evaluation pipeline; initial calibration and interpretability assessment.

Year 2: robust fusion (gating, missing-modality training); uncertainty estimation; first major publication.

Year 3: constrained agent design; offline RL/simulation; workflow-centred evaluation; second major publication.

Year 4: external validation and monitoring plan; prototype hardening for research deployment; thesis writing and final publications.

7. 7. Expected Contributions

1. Robust multimodal fusion methods for cardiology that tolerate missing modalities and domain shift.
2. Calibrated, uncertainty-aware risk stratification with clinically aligned thresholds and escalation rules.
3. A constrained agentic assistance framework with auditable safety controls and human-in-the-loop governance.
4. An evaluation protocol linking model performance to workflow outcomes, suitable for translational medical AI research.

8. 8. Ethics, Privacy, and Responsible Research

All work will follow institutional ethics review processes, protect patient privacy through de-identification and secure access controls, and document model limitations and failure modes. The agentic module will remain assistive, not autonomous, and will require clinician confirmation. Monitoring and audit logs will be maintained to support accountability and safe iteration.

Applicant Signature: Amjed [Signature] Date: 01/07/2026