Required Structure (Use These Headings)

1. **Abstract** (150–200 words): problem, data, methods, key results, clinical takeaways.

2. **Background & Clinical Need** (0.5–1 page): who is affected, current workflow, pressing pain point.

3. **Data & Cohorts** (Week 2): sources, inclusion/exclusion; exploratory analysis, cleaning, missingness; radiomics pipeline. Include one figure (data flow) and one table (cohorts).

4. **Methods:**
   • **Classical ML (Week 3):** LR/RF/KNN, SMOTE; splits; hyperparams.
   • **DL & Segmentation (Week 4):** pretrained CNN(s) and one segmentation model; fine-tuning details.
   • **Fairness & Mitigation (Week 5):** sensitive attributes; mitigation (oversampling/augmentation or others).
   • **Multi-Modal & Foundation (Week 6):** early concat, cross-attention; MedCLIP adaptation (linear probe, partial FT, LoRA); report trainable params.
   • **Robustness/Affordability (Week 8 lens):** order/quantity invariance tests (if LLM/VLM), input perturbations, external/site shift (if available); latency, cost/exam, energy/study.

5. **Evaluation Protocols & Metrics** (define before results):
   • **Performance:** AUROC, Accuracy, F1, AUPRC, sensitivity/specificity; segmentation Dice/IoU.
   • **Fairness:** per-site/subgroup sensitivity/specificity, gap metrics (max drop Δ), calibration by subgroup (ECE/Brier).
   • **Explainability:** Grad-CAM/saliency for imaging; SHAP/feature importances for tabular; text rationales if used.
   • **Ops metrics:** P95 latency (CPU/GPU), trainable parameters, cost/exam, energy/study (Wh), if available.

6. **Results** (figures + tables):
   • **ML vs. DL vs. MM/FM comparison table** (see Template A).
   • ROC/PR curves; confusion matrices; Dice/IoU; fairness disaggregations; calibration plots.
   • Explainability panels (e.g., Grad-CAM grid).

7. **Developed AI Application** (2–3 pages): use case developed through Streamlit, assistive scope, integration point, invariance & guideline checks, offline/edge constraints, shadow-mode edits/time saved, go/no-go criteria.

8. **Discussion & Limitations:** clinical relevance, generalizability, risks, algorithmovigilance plan (drift, rollback).

9. **CLAIM & FUTURE-AI Reporting:** completed checklists + commentary on any N/A items.

10. **Self-Reflection** (1–2 pages): technical growth, failures turned into design changes, ethics/robustness insights, help needed.

Tie-back to Weekly Deliverables (must be traceable)

Week 2 (EDA, cleaning, radiomics, PCA/t-SNE) → *Data & Cohorts*
Week 3 (splits, LR/RF/KNN, SMOTE, ROC/confusion) → *ML Methods/Results*
Week 4 (CNN fine-tune, segmentation, CSV metrics) → *DL Methods/Results*

So far, the previous weeks have been submitted in the MidTerm, so you need to compile a full report with the previous one as well for completeness purposes.

Week 5 (fairness mitigation + re-eval) → *Fairness*
Week 6 (concat, cross-attn, MedCLIP probing/LoRA, cost) → *MM/FM*

Template A — Model Comparison (include in Results)

Model	Type	Precision	Recall	AUC	F1	ECE
LogReg (radiomics)	ML	—	—			
RF (radiomics)	ML	—	—			
CNN FT #1	DL					
MM: Concat	MM					
MM: Cross-Attn	MM					
FM: MedCLIP (LP)	FM					
FM: MedCLIP (LoRA)	FM					

Template B — Fairness/Explainability Summary

Subgroup/Site	Sens.	Spec.	ECE	Notes (mitigation)
Site A / Subgroup 1				
Site B / Subgroup 2				

CLAIM & FUTURE-AI Checklist (fill and include)

Item (condensed)	Done?	(section/fig)
Clinical need, intended use, target population	[ ]	
Data sources, inclusion/exclusion, missingness	[ ]	
Reference standard/labels + uncertainty	[ ]	
Train/val/test split, leakage prevention	[ ]	
Model description (ML/DL/MM/FM), hyperparams	[ ]	
Evaluation metrics (perf+calibration)	[ ]	
Robustness tests (perturb., shift, invariance)	[ ]	
Fairness (subgroup/site analysis)	[ ]	
Explainability (images/tabular/text)	[ ]	
External validation / site shift (if any)	[ ]	
Deployment context (assistive scope, guardrails)	[ ]	
Ops metrics (latency, cost, energy)	[ ]	
Limitations, failure analysis, incidents plan	[ ]	
Reproducibility (code/data access)	[ ]	

Self-Reflection Prompts (address explicitly)

- What changed in your problem framing from Week 2 to Week 8? Which failure(s) taught you the most?
- Where did fairness or calibration gaps appear, and how did you respond?
- If you had to deploy in a low-resource site next month, what would you change (models, pipeline, guardrails)?
- What support do you need (IRB, hospital partner, FL orchestration, metric dashboards)?