Guideline Item	Status	Location in Report
Clinical need and intended use clearly defined	✓ Complete	Section 1 (Background)
Target population and clinical setting described	✓ Complete	Sections 1, 6
Data sources and acquisition methods documented	✓ Complete	Section 2 (Data & Cohorts)
Inclusion/exclusion criteria transparent	✓ Complete	Section 2
Handling of missing data explained	✓ Complete	Section 3.1 (SimpleImputer with median imputation)
Reference standard and label generation process	✓ Complete	Section 2, 7.4 (NLP-derived labels with ~10% noise)
Train/validation/test split clearly described	✓ Complete	Section 2 (patient-level separation)
Prevention of data leakage	✓ Complete	Section 3.1 (SMOTE within CV folds, not before)
Model architecture and hyperparameters	✓ Complete	Section 3 (all subsections with detailed configs)
Training procedures detailed	✓ Complete	Section 3 (including failed and successful approaches)
Evaluation metrics predefined	✓ Complete	Section 4 (AUROC, AUPRC, F1, precision, recall)
Performance metrics for all models	✓ Complete	Section 5.1 (comprehensive comparison table)
Calibration analysis	✓ Complete	Section 5.6 (ECE by subgroup for FM model)
Robustness testing	✓ Complete	Section 5.5 (Gaussian noise perturbation testing)

Fairness analysis across subgroups	✓ Complete	Section 5.3 (detailed analysis by AP/PA view type)
Fairness mitigation strategies	✓ Complete	Section 5.3 (SMOTE for ML; loss re-weighting attempted for DL with documented failure)
Explainability methods applied	✓ Complete	Section 5.4 (Grad-CAM for DL, SHAP for ML)
External validation	x Not done	Acknowledged as critical limitation (Sec 7.4)
Comparison with clinical standard of care	x Not done	No radiologist comparison (noted in Sec 7.4)
Clinical deployment context specified	✓ Complete	Section 6 (assistive tool with detailed use cases)
Assistive vs autonomous scope clarified	✓ Complete	Section 6.1 (explicitly assistive, never autonomous)
Guardrails and safety mechanisms	✓ Complete	Section 6.1 & 6.4 (notification-only, subgroup monitoring, user control)
Operational metrics (latency, cost)	✓ Complete	Section 5.5 (P95 latency, cost per 1k images)
Energy efficiency considerations	✓ Complete	Section 5.5 (Wh per image for all models)
Limitations clearly stated	✓ Complete	Section 7.4 (10 detailed limitations)
Failure modes analyzed	✓ Complete	Sections 5.2, 5.3, 7.2, 7.3, 9 (MedCLIP failure, mitigation failure)
Monitoring and incident response plan	✓ Complete	Section 7.6 (algorithmovigilance with triggers)
Code and data availability	✓ Complete	Code available (Jupyter notebook); data is public MIMIC-CXR