



No	Item 	Reported 	Location (if Yes/Partial)
1. Model Specification			
1.1	Model name, vendor/developer, version/identifier, release date, and training/knowledge cutoff date	☐ Yes ☒ Partial ☐ No ☐ N/A	Methods
1.2	Model architecture and key characteristics	☐ Yes ☐ Partial ☒ No ☐ N/A	Page/Section
1.3	Model pretraining, post-training, and inference-time adaptation strategy	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
1.4	Modality support details (input and output) and limitations	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods, D
1.5	Language capabilities	☐ Yes ☐ Partial ☒ No ☐ N/A	Page/Section
1.6	Model access	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
1.7	Sharing of code, data, and model artifacts	☐ Yes ☒ Partial ☐ No ☐ N/A	Supplement
1.8	Computational requirements	☐ Yes ☐ Partial ☐ No ☒ N/A	Page/Section
2. Prompt Design			
2.1	Prompt engineering protocol with versioning	☐ Yes ☐ Partial ☒ No ☐ N/A	Page/Section

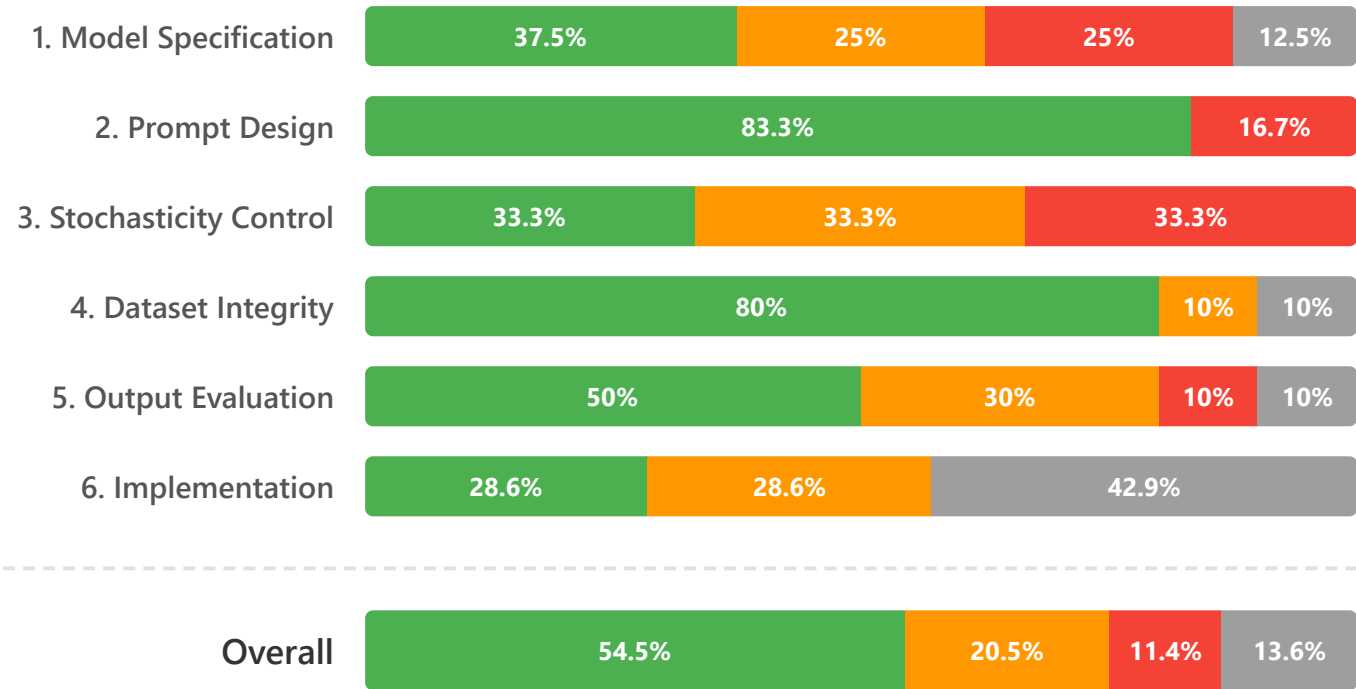
No	Item 	Reported 	Location (if Yes/Partial)
2.2	Prompting strategy, format, and length	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
2.3	Prompt modality, language, technical input specification, and full content	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
2.4	Integration of relevant patient clinical context	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
2.5	Interaction style and session memory policy	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
2.6	Output handling	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
3. Stochasticity Control			
3.1	Generation parameters	☐ Yes ☐ Partial ☒ No ☐ N/A	Page/Section
3.2	Prompt operator characteristics and number of prompt attempts	☐ Yes ☒ Partial ☐ No ☐ N/A	Methods
3.3	Output selection	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
4. Dataset Integrity			
4.1	Dataset name, version, access type, source citation, license, and compliance statement	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
4.2	Dataset origin	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
4.3	Ethics and consent statements	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods

No	Item 	Reported 	Location (if Yes/Partial)
4.4	Prior dataset usage and publication date	☐ Yes ☒ Partial ☐ No ☐ N/A	Methods, D
4.5	Dataset composition and data synthesis details	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods, R
4.6	Sample characteristics and representational bias analysis	☒ Yes ☐ Partial ☐ No ☐ N/A	Results (Ta
4.7	Reference standard and annotator qualifications	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
4.8	Preprocessing and data pairing/registration	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods
4.9	Missing data extent, mechanism, and handling	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods; F
4.10	Separation of training, fine-tuning, internal testing, and external testing datasets	☐ Yes ☐ Partial ☐ No ☒ N/A	Page/Section
5. Output Evaluation			
5.1	Output evaluation method and performance metrics	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods (C
5.2	Human evaluator characteristics and reliability analysis	☐ Yes ☒ Partial ☐ No ☐ N/A	Methods, D
5.3	Statistical analysis of evaluation results	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods (S
5.4	Subgroup performance and output bias assessment	☒ Yes ☐ Partial ☐ No ☐ N/A	Results
5.5	Failure analysis and error metrics	☐ Yes ☒ Partial ☐ No ☐ N/A	Results (Ta

No	Item 	Reported 	Location (if Yes/Partial)
5.6	Output stochasticity and reproducibility constraints	☒ Yes ☐ Partial ☐ No ☐ N/A	Methods (T
5.7	Performance effects of different prompt strategies and revisions	☐ Yes ☐ Partial ☒ No ☐ N/A	Page/Section
5.8	Model version comparisons and temporal performance variation	☐ Yes ☒ Partial ☐ No ☐ N/A	Methods (T
5.9	Methods for explainability and interpretability of model outputs	☐ Yes ☐ Partial ☐ No ☒ N/A	Page/Section
5.10	Comparison with clinically relevant benchmarks	☒ Yes ☐ Partial ☐ No ☐ N/A	Results, Di
6. Implementation			
6.1	Declared intended application and scope of use	☒ Yes ☐ Partial ☐ No ☐ N/A	Introduction
6.2	Clinical workflow integration	☐ Yes ☐ Partial ☐ No ☒ N/A	Page/Section
6.3	Measured clinical utility or added value	☐ Yes ☒ Partial ☐ No ☐ N/A	Discussion
6.4	Model limitations, explicit clinical non-use cases, and potential misuse considerations	☒ Yes ☐ Partial ☐ No ☐ N/A	Discussion
6.5	Safety testing and monitoring protocols	☐ Yes ☐ Partial ☐ No ☒ N/A	Page/Section
6.6	Data security and privacy safeguards	☐ Yes ☒ Partial ☐ No ☐ N/A	Methods
6.7	Governance, auditability, and oversight	☐ Yes ☐ Partial ☐ No ☒ N/A	Page/Section

Summary of Checklist Completion by Section

Yes Partial No N/A Empty



*Percentages are calculated independently and may not sum to exactly 100% due to rounding.

Note: Please cite the following article when using this checklist: Mese I, Akinci D'Antonoli T, Bluethgen C, Bressem K, Cuocolo R, Chaudhari A, Tejani AS, Isaac A, Ponsiglione A, Meddeb A, Khosravi B, Le Guellec B, Kahn CE Jr, Suh CH, Pinto Dos Santos D, Koh DM, Tzanis E, Kotter E, Colak E, Kitamura F, Busch F, Nensa F, Yang G, Müller H, Kather JN, Nawabi J, Kleesiek J, Zhong J, Santinha J, Haubold J, de Almeida JG, Lekadir K, Marias K, Reiner LN, Maier-Hein L, Moy L, Adams LC, Martí-Bonmatí L, Paschali M, Moassefi M, Dietzel M, Huisman M, Ingrisich M, Klontzas ME, Papanikolaou N, Diaz O, Kuriki P, Seeböck P, Rouzrokh P, Strotzer QD, Park SH, Faghani S, Tayebi Arasteh S, Kim SH, Venugopal VK, Kim W, Kocak B. Reporting checklist for foundation and large language models in medical research (REFINE): an international consensus guideline. *Diagn Interv Radiol.* 2026 Feb 26. doi: 10.4274/dir.2026.263812. Epub ahead of print. PMID: 41742713.