**MI-CLAIM checklist for generative AI clinical studies**

|  |  |  |
| --- | --- | --- |
| **Before paper submission** | | |
| **Study design (Part 1)** | **Page number** | **Notes if not completed** |
| The clinical problem in which the model will be employed is clearly detailed in the paper. |  |  |
| The research question is clearly stated. |  |  |
| All cohort selection criteria and study design are detailed in such a way that they can be reproduced by an external researcher. |  |  |
| Details on how labels were generated are described, including any annotation guidelines, level of experience of annotators, inter-annotator scores, etc. |  |  |
| Is the output data type categorical, continuous, or unstructured? | ☐ Categorical  ☐ Continuous  ☐ Unstructured | |
| The characteristics of the cohorts are detailed in the text and are shown to be representative of real-world clinical settings. |  |  |
| **Resources and optimization (Part 2)** | **Page number** | **Notes if not completed** |
| Model/application components are clearly detailed including: base model(s) used, embedding model(s), retrieval model(s), and other auxiliary models or tools. |  |  |
| The origin of all data sources for model training, finetuning, or inference is described and the original format is detailed in the paper. |  |  |
| All data preprocessing for model training, finetuning, or inference is described, including appropriate randomization and other transformations. |  |  |
| The independence between training, validation (including for prompt engineering), and test sets has been described, and data is split at the patient level. |  |  |
| **Model performance and evaluation (Parts 3-4)** | **Page number** | **Notes if not completed** |
| The state-of-the-art solution used as a baseline for comparison has been identified and detailed. |  |  |
| The performance comparison between the baseline and the proposed model is presented with the appropriate statistical significance. |  |  |
| Identify what evaluation(s) were performed, and provide clear justifications for the primary metrics used for each evaluation. | ☐Overlap accuracy  ☐Semantic accuracy  ☐Clinical utility |  |
| If applicable, details on human evaluation are described, including any evaluation guidelines, level of experience of evaluators, inter-reviewer scores, etc. |  |  |
| **Model examination (Part 5)** | **Page number** | **Notes if not completed** |
| Relevant interpretability techniques, error analysis, and/or other approaches are applied to demonstrate an absence of unreasonable risk and brittleness, including a low risk of catastrophic and especially undetected failure. |  |  |
| A discussion of the risk revealed by the examination results is presented with respect to model/algorithm performance. |  |  |
| Which step(s) have been taken to understand model biases, privacy and security concerns, and other potential harm? | ☐ Discussion  ☐ Identification  ☐ Mitigation |  |
| A discussion and/or assessment of relevant distribution shifts and their impact on the model's performance has been provided |  |  |
| The authors provide recommendations or discussion of post-deployment evaluation |  |  |
| **Reproducibility (Part 6)** | **Page number** | **Notes** |
| **Data transparency: choose appropriate tier of transparency** | | |
| Tier 1: complete sharing of the code and data, including all prompts tested, software dependencies, and evaluation setups. |  |  |
| Tier 2A: complete sharing of the code with synthetic data provided |  |  |
| Tier 2B: complete sharing of the code |  |  |
| Tier 3: no sharing of code or data |  |  |
| **Model transparency** | | |
| Model hyperparameters, along with infrastructure and compute requirements for running and/or developing the model are included, specifying hardware type, costs, and training time where applicable. |  |  |
| A clinical model card is included summarizing the model capabilities, intended use, descriptions of any dataset or other integrations, limitations, potential biases, and risks |  |  |
| If applicable: Model weights are released to a secure repository and/or with appropriate use agreements. |  |  |