

Identify the study as developing, fine-tuning and/or evaluating the performance of an LLM, specifying the task, the target population and the outcome to be predicted.

Author

August 2026

1	Abstract
2	Introduction
3	Methods
3.1	Data
3.2	Analytical methods
3.3	LLM Output
3.4	Annotation
3.5	Prompting
3.6	Summarization
3.7	Instruction tuning/alignment
3.8	Compute
3.9	Ethical approval
3.10	Open science
3.11	Public involvement
4	Results
4.1	Participants
4.2	Performance
4.3	LLM updating
5	Discussion
5.1	Interpretation
5.2	Limitations
5.3	Usability of the LLM in context
6	Appendix

Table 1: For studies using existing LLMs, users should include reference(s) to reportable information if provided by the original developers or state that this information is not available. M, LLM methods; D, de novo LLM development; E, LLM evaluation; H, LLM evaluation in healthcare settings; C, classification; OF, outcome forecasting; QA, long-form question answering; IR, information retrieval; DG, document generation; SS, summarization and simplification; MT, machine translation; API, application programming interface.

Section	Item	Description	Research Design	LLM task	Link
Title	1	Identify the study as developing, fine-tuning and/or evaluating the performance of an LLM, specifying the task, the target population and the outcome to be predicted.	All	All	
Abstract	2	See TRIPOD-LLM for abstracts.	All	All	
Introduction					
Background	3a	Explain the healthcare context/use case (for example, administrative, diagnostic, therapeutic and clinical workflow) and rationale for developing or evaluating the LLM, including references to existing approaches and models.	All	All	2
	3b	Describe the target population and the intended use of the LLM in the context of the care pathway, including its intended users in current gold standard practices.	E/H	All	2

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Objectives	4	Specify the study objectives, including whether the study describes the initial development, fine-tuning or validation of an LLM (or multiple stages).	All	All	2
Methods					
Data	5a	Describe the sources of data separately for the training, tuning and/or evaluation datasets and the rationale for using these data (for example, web corpora, clinical research/trial data, EHR data or unknown).	All	All	3.1
	5b	Describe the relevant data points and provide a quantitative and qualitative description of their distribution and other relevant descriptors of the dataset.	All	All	3.1
	5c	Specifically state the date of the oldest and newest item of text used in the development process (training, fine-tuning and reward modeling) and the evaluation datasets.	All	All	3.1
	5d	Describe any data pre-processing and quality checking, including whether this was similar across text corpora, institutions and relevant sociodemographic groups.	All	All	3.1

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Analytical Methods	5e	Describe how missing and imbalanced data were handled and provide reasons for omitting any data.	All	All	3.1
	6a	Report the LLM name, version and last date of training.	All	All	3.2
	6b	Report details of the LLM development process, such as LLM architecture, training, fine-tuning procedures and alignment strategy and alignment goals.	M/D	All	3.2
	6c	Report details of how the text was generated using the LLM, including any prompt engineering and inference settings (for example, seed, temperature, max token length and penalties).	M/D/E	All	3.2
	6d	Specify the initial and postprocessed output of the LLM (for example, probabilities, classification and unstructured text).	All	All	3.2
	6e	Provide details and rationale for any classification, and if applicable, how the probabilities were determined and thresholds identified.	All	C/OF	3.2

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
LLM Output	7a	Include metrics that capture the quality of generative outputs, such as consistency, relevance, accuracy and presence/type of errors compared to gold standards.	All	QA, IR, DG, SS, & MT	3.3
	7b	Report the outcome metrics’ relevance to the downstream task at deployment time and, where applicable, the correlation of metric to human evaluation of the text.	E/H	All	3.3
	7c	Clearly define the outcome, how the LLM predictions were calculated, the date of inference for closed-source LLMs and evaluation metrics.	E/H	All	3.3
	7d	If outcome assessment requires subjective interpretation, describe the qualifications of the assessors, any instructions provided, relevant information on demographics of the assessors and interassessor agreement.	All	All	3.3
	7e	Specify how performance was compared to other LLMs, humans and other benchmarks or standards.	All	All	3.3

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Annotation	8a	If annotation was done, report how the text was labeled, including providing specific annotation guidelines with examples.	All	All	3.4
	8b	If annotation was done, report how many annotators labeled the dataset(s), including the proportion of data in each dataset that was annotated by more than one annotator, and the interannotator agreement.	All	All	3.4
	8c	If annotation was done, provide information on the background and experience of the annotators or the characteristics of any models involved in labeling.	All	All	3.4
Prompting	9a	If research involved prompting LLMs, provide details on the processes used during prompt design, curation and selection.	All	All	3.5
	9b	If research involved prompting LLMs, report what data were used to develop the prompts.	All	All	3.5
Summarization	10	Describe any preprocessing of the data before summarization.	All	SS	3.6

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Instruction tuning/alignment	11	If instruction tuning/alignment strategies were used, what were the instructions, data and interface used for evaluation, and what were the characteristics of the populations doing the evaluation?	M/D	All	3.7
Compute	12	Report compute, or proxies thereof (for example, time on what and how many machines, cost on what and how many machines, inference time, floating-point operations per second).	M/D/E	All	3.8
Ethical approval	ap-13	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent.	All	All	3.9
Open science	14a	Give the source of funding and the role of the funders for the present study.	All	All	3.10
	14b	Declare any conflicts of interest and financial disclosures for all authors.	All	All	3.10
	14c	Indicate where the study protocol can be accessed or state that a protocol was not prepared.	H	All	3.10

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
	14d	Provide registration information for the study, including register name and registration number, or state that the study was not registered.	H	All	3.10
	14e	Provide details of the availability of the study data.	All	All	3.10
	14f	Provide details of the availability of the code to reproduce the study results.	All	All	3.10
Public involvement	15	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation or dissemination of the study or state no involvement.	H	All	3.11
Results					
Participants	16a	When using patient/EHR data, describe the flow of text/EHR/patient data through the study, including the number of documents/questions/participants with and without the outcome/label.	E/H	All	4.1

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
	16b	When using patient/EHR data, report the characteristics overall and for each data source or setting and development/evaluation splits, including the key dates, key characteristics and sample size.	E/H	All	4.1
	16c	For LLM evaluation that includes clinical outcomes, show a comparison of the distribution of important clinical variables that may be associated with the outcome between development and evaluation data.	E/H	All	4.1
	16d	When using patient/EHR data, specify the number of participants and outcome events in each analysis.	E/H	All	4.1
Performance	17	Report LLM performance according to prespecified metrics and/or human evaluation.	All	All	4.2
LLM updating	18	If applicable, report the results from any LLM updating, including the updated LLM and subsequent performance.	All	All	4.3
Discussion					

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Interpretation	19a	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies.	All	All	5.1
Limitations	19b	Discuss any limitations of the study and their effects on any biases, statistical uncertainty and generalizability.	All	All	5.2
Usability of the LLM in context	19c	Describe any known challenges in using data for the specified task and domain context with reference to representation, missingness, harmonization and bias.	E/H	All	5.3
	19d	Define the intended use for the implementation under evaluation, including the intended input, end-user and level of autonomy/human oversight.	E/H	All	5.3
	19e	If applicable, describe how poor quality or unavailable input data should be assessed and handled when implementing the LLM.	E/H	All	5.3
	19f	If applicable, specify whether users will be required to interact in the handling of the input data or use of the LLM, and what level of expertise is required of users.	E/H	All	5.3

Continued on next page

Table 1 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
	19g	Discuss any next steps for future research, with a specific view of the applicability and generalizability of the LLM.	All	All	5.3

Table 2: TRIPOD-LLM for Abstracts. RLHF, reinforcement learning with human feedback. For studies using existing LLMs, users should include reference(s) to reportable information if provided by the original developers or state that this information is not available. M, LLM methods; D, de novo LLM development; E, LLM evaluation; H, LLM evaluation in healthcare settings; C, classification; OF, outcome forecasting; QA, long-form question answering; IR, information retrieval; DG, document generation; SS, summarization and simplification; MT, machine translation; API, application programming interface.

Section	Item	Description	Research Design	LLM task	Link
Title	2a	Identify the study as developing, fine-tuning and/or evaluating the performance of an LLM, specifying the task, the target population and the outcome to be predicted.	All	All	1
Background	2b	Provide a brief explanation of the healthcare context, use case and rationale for developing or evaluating the performance of an LLM.	E & H	All	1
Objectives	2c	Specify the study objectives, including whether the study describes LLMs development, tuning and/or evaluation	All	All	1

Continued on next page

Table 2 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Methods	2d	Describe the key elements of the study setting.	All	All	1
	2e	Detail all data used in the study, specify data splits and any selective use of data.	M, D, & E	All	1
	2f	Specify the name and version of LLM(s) used.	All	All	1
	2g	Briefly summarize the LLM-building steps, including any fine-tuning, reward modeling and RLHF.	M & D	All	1
	2h	Describe the specific tasks performed by the LLMs (for example, medical QA, summarization and extraction), highlighting key inputs and outputs used in the final LLM.	All	All	1
	2i	Specify the evaluation datasets/populations used, including the endpoint evaluated, and detail whether this information was held out during training/tuning where relevant and what measure(s) were used to evaluate LLM performance.	All	All	1
Results	2j	Give an overall report and interpretation of the main results.	All	All	1

Continued on next page

Table 2 – continued from previous page

Section	Item	Description	Research Design	LLM task	Link
Objectives	2k	Explicitly state any broader implications or concerns that have arisen in light of these results.	All	All	1
Other	2c	Give the registration number and name of the registry or repository (if relevant).	H	All	1