Online Resource 2: CHART checklist

Nobody writes for the reader: reading level in paediatric health guidance, real health queries and AI chatbot responses

European Journal of Pediatrics

Conal Cunningham, Colm Breatnach

Corresponding author: Conal Cunningham, Paediatric Intensive Care Unit, Children's Health Ireland at Crumlin, Dublin, Ireland, conal.cunningham2@childrenshealthireland.ie

Referenced from §2.1. The Chatbot Assessment Reporting Tool (CHART) statement (Ann Fam Med 2025;23(5):389-98, jointly published in The BMJ, Artificial Intelligence in Medicine, BJS, BMC Medicine, BMJ Medicine and JAMA Network Open) specifies 12 items and 39 subitems. Item wording below is quoted from the published statement.

Item	Requirement	Status	Location
1a	State that the study is assessing one or more generative AI-driven chatbots for clinical evidence or health advice	Reported	Title
1b	Apply a structured format, if applicable	Reported	Abstract (Purpose/Methods/Results/Conclusion)
2a	State the scientific background, rationale, and healthcare context, referencing relevant literature	Reported	§1
2b	State the aims and research questions including target audience, intervention, comparator(s), and outcome(s)	Reported	§1, final paragraph
3a	State the name and version identifier(s) of the model(s)/chatbot(s), including date of release or last update	Partial	Model names and versions in Tables 1–2; version identifiers checked at start and close of collection (§2.4). Release/last-update dates are not stated.
3b	State whether open-source or closed-source/proprietary	Reported	§2.4 (four proprietary models and one open-weight model, DeepSeek V4)
4a	State whether base, novel base, tuned, or fine-tuned model	Not applicable	All five are externally deployed commercial base models, not tuned/fine-tuned by this study
4b	If base model, cite its development in sufficient detail	Reported	Model names/versions identify the deployed system.
4c	If tuned/fine-tuned, describe pre/post-deployment data	Not applicable	No fine-tuning performed

	and parameters		
5a	Describe the evolution of study prompt development	Reported	Adaptation sub-study arm instructions given verbatim (§2.5); main-corpus queries are patient-authored and selected as described in §2.3, not engineered.
5a(i)	Sources of prompts	Reported	§2.3 (Reddit forum posts, PushShift archive)
5a(ii)	Number/characteristics of individuals involved in prompt engineering	Not applicable	Queries are patient-authored, not investigator-engineered; the adaptation sub-study instructions are fixed, pre-specified strings (§2.5).
5a(iii)	Patient and public involvement in prompt engineering	Not applicable	No PPI process; queries are themselves patient-authored (this is the study's point of contrast with investigator-written-prompt studies, §1)
5b	Provide study prompts	**Partial by design**	Adaptation sub-study instructions given verbatim (§2.5). The 1,000 parent questions are deliberately not reproduced: Statements and Declarations (Data availability) explains why verbatim question text is withheld, and §2.8 states no verbatim post text is reproduced.
6a	Route of access to generative AI model	Reported	§2.4 (consumer web interfaces, not developer APIs, explicitly justified)
6b	Date(s) and location(s) of queries	Reported	§2.4 (9-16 June 2026, Dublin, Ireland)
6c	Whether prompts were input into separate chat sessions	Reported	§2.4 ("each in a fresh session")
6d	Provide all chatbot output/responses	Reported via repository	Statements and Declarations (Data availability): all 6,500 model responses are deposited at Zenodo (DOI 10.5281/zenodo.22846367), CC BY 4.0.
7a	Define the ground truth/reference standard for successful performance	Reported	§2.7 — the AMA sixth-grade and NIH eighth-grade recommendations are the explicit reference standards (this study evaluates readability, not clinical accuracy, so "ground truth" here is the readability standard, not a correctness criterion)
7b	Describe the process for performance evaluation	Reported	§2.6 (seven metrics, textstat library, computation described)
7b(i)	Number/characteristics of team members involved in evaluation	Not applicable	Primary outcome is automated scoring (textstat), not human evaluation.
7b(ii)	PPI during evaluation	Not applicable	No PPI process
7b(iii)	Whether evaluators were blinded to chatbot identity	Not applicable for automated scoring	Readability scoring (textstat) is a deterministic function of the text alone and cannot be influenced by chatbot identity.
8	Report how sample size was determined	Reported	1,000 queries, stratified by subspecialty (§2.3); 100-query adaptation subsample (§2.5); 50-query within-model reproducibility subsample (§2.4).
9a	Statistical analysis methods, including any evaluation of	Reported	Statistical methods §2.7; within-model reproducibility method §2.4.

	reproducibility		
9a(i)	Measures used for performance evaluation	Reported	§2.6 (seven named metrics)
10a	Report the performance evaluation, alignment with reference standard	Reported	§3.1, Table 1, throughout §3
10b	For responses deviating from the reference standard, state the nature of the difference	Reported	§3.2-3.4 (between-model variation, adaptation gap characterised throughout)
10c	Report evaluation for potentially harmful, biased, or misleading responses	Partial	No systematic harm/bias/misinformation evaluation was in scope for this readability-focused study
11a	Interpret findings in context of relevant evidence	Reported	§4.3, §4.5
11b	Strengths and limitations	Reported	§5 (Limitations)
11c	Implications for practice, education, policy, regulation, research	Reported	§4.4
12a	Conflicts of interest for all authors	Reported	Statements and Declarations (Competing interests)
12b	Sources of funding and their role	Reported	Statements and Declarations: "No funding was received for this work"
12c	Process undertaken for ethical approval	Reported	§2.8; Statements and Declarations (Ethics approval)
12c(i)	Data privacy safeguards	Reported	§2.8
12c(ii)	Permission/licensing for copyrighted data	Reported	Statements and Declarations (Data availability): guideline page text is not redistributed as it is third-party copyright; source URLs are released instead.
12d	Study protocol	Not provided	
12e	Where data, code, and model parameters can be accessed	Reported	Statements and Declarations (Data availability; Code availability): Zenodo, DOI 10.5281/zenodo.22846367.