DECIDE-AI checklist mapping

*Companion to "OpenConsult, an Open-Source Fully Local AI Scribe With Live Decision Support: Development and Formative Evaluation". DECIDE-AI [1] is written for the early **live** clinical evaluation of AI-based decision support: real users, real settings, small scale. This study is a synthetic-only, pre-deployment formative evaluation and therefore sits one rung below the guideline's intended stage. As set out in the Methods of the manuscript (How it was judged), we report against every item: **R** = reported (with location), **P** = partially reported (with what is and is not covered), **N/A** = cannot apply at this stage (with the reason stated).*

Table S1. AI-specific items.

#	Item	Status	Where / why
1	Title identifies early clinical evaluation of an AI-based decision support system	**P**	The title identifies the system as an AI scribe with live decision support and the study as a formative evaluation. The DECIDE-AI positioning, one stage before live clinical evaluation, is stated in the Abstract and Methods (How it was judged), not in the title.
2a	Clinical problem, current practice, intended patient population	**R**	Introduction (documentation burden, the ambient AI landscape, UK policy, regulation and cost). Methods (OpenConsult) states that the system is for research and education only and is not for use in real patient care; it was built and evaluated in general practice.
2b	Intended users, integration in care pathway, anticipated impact	**R**	Methods (OpenConsult): intended use. Methods (A consultation with OpenConsult) gives the intended consultation flow: the clinician is the sole decision-maker, and nothing becomes part of the record until the doctor has read and approved it. Figures 2-8.
3a	Patient recruitment, eligibility, sample size	**N/A**	No patients: every consultation processed has been synthetic, either scripted or acted (Methods (How it was judged); Methods (Ethical Considerations)). The synthetic case mix is reported instead (Results, Table 1; Multimedia Appendix 4).
3b	User recruitment and number	**N/A**	No users beyond the clinician-developer. This is stated in Discussion (Limitations) and under Conflicts of Interest; independent clinician raters are named as future work.
3c	User training and familiarisation	**N/A**	The single user built the system (Methods (How it was built)). No other user has been trained, and

			no claim is made about use by others.
4a	AI system: version, algorithm type, development population, preclinical performance	**R**	Methods (OpenConsult) names the components and models (MedGemma 27B served through Ollama, a fast Whisper implementation, WhisperX with pyannote, a local embedding model with a PostgreSQL vector store, a local speech engine). Exact versions and settings are in Multimedia Appendix 2 and in the lockfile of the archived release. Formative performance is in Results and Table 1.
4b	Input data, acquisition, preprocessing, missing/low-quality data handling	**R**	Methods (A consultation with OpenConsult): room audio is transcribed live, and the recording is re-processed with speaker separation at the end. Uncertain load-bearing values are flagged, and the system refuses to draft from audio that never became usable text (Methods (A consultation with OpenConsult); Figure 5). Thresholds are in Multimedia Appendix 2.
4c	Outputs and their presentation to users	**R**	Methods (A consultation with OpenConsult); Figures 2-8 (differential grading, urgency alarm, cited note, audit log); Multimedia Appendix 3 (raw-segment view, referral letter).
5a	Evaluation setting(s)	**R**	Methods (How it was judged): automated tests on a fixed bench of scripted consultations, plus acted consultations in the consulting room on the development machine. The capture method for the worked example is stated in Results.
5b	Workflow integration, timing, decision-making process	**R**	Methods (A consultation with OpenConsult) and Figures 2-8: live support during the encounter, note and letter after the patient has left, and an approval gate before anything is final.
6a	How significant errors/malfunctions were identified (methods)	**R**	Methods (How it was judged): checks built alongside each feature with marking schemes written before any run, gates that refuse on failure, acted consultations as a second modality, and an adversarial security review whose provenance is stated.
6b	Risk identification, analysis, mitigation	**R**	Methods (A consultation with OpenConsult), final paragraph (three tiers of structural protection, each with its limit); Methods (How it was judged) (threat model and scope of the security review).
7	Human factors methods	**N/A**	No formal human-factors or usability assessment exists beyond the developer's own use, and none

			is claimed. The Discussion sets out research questions on clinicians' reviewing behaviour that such work would address.
8	Ethics-related methods	**P**	Methods (Ethical Considerations): synthetic-only by design; the Health Research Authority decision tools classed the work as research requiring no NHS Research Ethics Committee review; the volunteers who role-played patients gave written consent; their recordings are held locally as personal data. No further ethics framework was applied because no patient-facing research was conducted.
9a	Patient characteristics and missing data (results)	**N/A**	No patients. The synthetic case mix is characterised in Table 1 and Multimedia Appendix 4.
9b	User characteristics (results)	**R**	Methods (How it was built) and Discussion (Limitations): one clinician-developer, a practising general practitioner, who also scripted and graded the evaluation.
10a	Exposure, usage, adherence (results)	**P**	The development record (213 commits over 27 days; Results), the five acted consultations (Table 2) and the worked example (Figures 2-8) report actual usage. No adherence measure applies with a single developer-user.
10b	Changes to clinical workflow caused by the system	**N/A**	No deployment into real clinical workflow has occurred.
11	Modifications to the system during the study	**R**	Table 2 (the fix for each defect, each with its regression test); Methods (How it was built) (one reviewable change at a time). The full modification history is under version control and published with the code.
12	Human–computer agreement	**N/A** (future work)	Not measured. Evaluation by independent clinician raters (Discussion, Future work) is the planned instrument for it.
13a	Significant errors: rates, causes, patient-care implications	**R**	Table 2 (defects found in acted consultations, with causes and why the test suite missed them); Table 3 (security findings by severity and disposition); the fabricated-note specimen and the gate it produced (Results).
13	Safety risks / observed harm	**R** (scope	No patient harm is possible at this stage because there are no patients. The realised risk found, a

b		-bounded)	stored cross-site-scripting vulnerability, is reported with its fix and verification (Results; Table 3).
14a	Usability evaluation against recognised standards	**N/A**	Not performed, and not claimed.
14b	User learning curves	**N/A**	Single developer-user; no learning-curve claim is made.
15	Whether results support the intended use	**R**	Discussion, including its Conclusions: the system is presented as a worked, auditable proof of concept and an open research instrument. It behaves as intended under formative, synthetic evaluation; no claim of clinical validity is made.
16	Safety profile and implications (discussion)	**R**	Discussion (Limitations; Future work, including a scribe-only edition aligned with the current regulatory position); Methods (OpenConsult) (regulatory status; not for use in patient care).
17	Data and code availability	**R**	Data Availability: source code public under AGPL-3.0 with archived DOIs; no patient dataset exists; evaluation material is synthetic and in the repository.

Table S2. Generic items.

#	Item	Status	Where / why
I	Structured abstract	**R**	Abstract (Background, Objective, Methods, Results, Conclusions).
II	Study objectives	**R**	Abstract (Objective); Introduction (Objective).
III	Protocol, registration, ethics approval	**P**	No prospective protocol or registration: this is a reported development, not a planned clinical study. Metrics and their definitions were committed to the repository before any data existed (Methods (How it was judged)), and the evaluation protocol for the fuller spoken mode is fixed in the repository before use (Methods (A consultation with OpenConsult)). No NHS ethics committee review was required (Methods (Ethical Considerations), with the reason).
IV	Primary and secondary outcomes	**P**	No trial outcomes. The per-script marking schemes and gate criteria (Table 1), fixed before the evaluation runs, are the closest formative equivalent.

V	Statistical analysis methods	**P**	Deterministic settings (temperature 0, fixed seed; Multimedia Appendix 2) and mechanical, code-enforced checks are described. No inferential statistics are used or claimed.
VI	Patient and public involvement	**N/A**	No patients. Public involvement begins with the open-source release; the Discussion invites independent evaluation.
VII	Main results for prespecified outcomes	**R**	Results, against the marking schemes and gates of Table 1.
VIII	Subgroup analyses	**N/A**	Not applicable to a formative study of a single system.
IX	Strengths and limitations	**R**	Discussion (Limitations).
X	Conflicts of interest and funding	**R**	Conflicts of Interest (self-built and self-evaluated status declared) and Funding (none).

Summary. Across the 38 rows of the checklist (27 items, several with sub-items): 21 are reported in full, 6 partially (each with what is missing stated), and 11 are not applicable at this pre-deployment, synthetic-only stage, each with its reason given above. The not-applicable items concentrate in patients, users and human factors. That is why the manuscript positions itself one rung below DECIDE-AI's intended stage and names the evaluation that would close the gap as future work.

References

1. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-933. doi:10.1038/s41591-022-01772-9 PMID:35585198