IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

Context-Aware Adaptive Retrieval-Augmented Generation for Evidence-Grounded Parkinson's Disease Clinical Literature Question Answering

Emerson B[1], Shanjana P[2]

M.Tech. Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai, Tamil Nadu, India[1,2]

Abstract: Parkinson's disease (PD) research spans clinical examination, speech and voice analysis, gait assessment, handwriting, imaging, biomarkers and longitudinal outcomes. Evidence relevant to these modalities is distributed across heterogeneous studies that differ in population, disease stage, acquisition protocol, outcome definition and study design. Retrieval-augmented generation (RAG) can connect language models to external literature, but retrieval relevance alone does not establish that a study is applicable to a particular question or context. This research proposal presents a context-aware adaptive RAG framework for evidence-grounded question answering over PD literature. The framework explicitly represents available, missing and unknown assessment modalities; retrieves candidate evidence; evaluates applicability across population, disease stage, modality, outcome and study design; decomposes responses into atomic claims; and links claims to supporting or conflicting passages. A response controller selects among qualified answering, further retrieval, clarification and abstention. A comparative evaluation protocol is proposed using LLM-only, standard RAG, hybrid RAG and generic adaptive-RAG baselines. No implementation or experimental evaluation has yet been performed; this manuscript therefore describes a proposed framework and evaluation plan rather than measured performance. The intended use is literature support and evidence navigation, not diagnosis or treatment recommendation.

Keywords: Parkinson's disease, retrieval-augmented generation, medical question answering, evidence applicability, adaptive retrieval, claim verification, uncertainty, abstention.

I. INTRODUCTION

Parkinson's disease is a clinically heterogeneous neurodegenerative disorder. Research and clinical assessment draw on motor examination, non-motor symptoms, speech and voice characteristics, gait, handwriting, neuroimaging, biomarkers and patient history. These evidence sources are not interchangeable. A result derived from a particular cohort, disease stage, acquisition protocol or outcome measure may not transfer directly to another population or assessment setting.

The growing volume of PD literature creates an evidence-navigation challenge. A clinician, researcher or student may need to identify relevant studies, compare findings, understand conflicting results and determine whether evidence applies to the context described. Conventional keyword search returns documents but leaves synthesis and source appraisal to the reader. Large language models can produce fluent summaries, yet may generate unsupported statements, omit caveats or cite sources that do not substantiate individual claims.

Retrieval-augmented generation combines information retrieval with language generation. In a basic RAG pipeline, documents are retrieved for a query and supplied to a language model as context. However, a high retrieval score does not guarantee clinical applicability. A paper about voice-based classification in a selected cohort may be topically relevant to a question about gait assessment but not directly applicable. Evidence from early-stage participants may not support a claim about advanced disease.

Recent work has explored uncertainty-guided retrieval and abstention in medical question answering, claim-level verification in biomedical RAG, and Parkinson's-focused conversational AI. Therefore, generic adaptive retrieval, claim verification, abstention and PD-focused RAG cannot individually be claimed as new contributions. This proposal focuses on explicit PD-context representation, evidence-applicability assessment and traceable claim-to-source mapping.

 This work is licensed under a Creative Commons Attribution 4.0 International License

II. AIMS AND PROPOSED CONTRIBUTIONS

An explicit context representation distinguishing available, missing and unknown assessment modalities without imputing absent information.
An evidence-applicability schema considering population, PD stage, modality, outcome and study design separately from topical relevance.
A traceable workflow linking atomic claims to supporting, conflicting, mixed or absent evidence passages.
A response policy covering qualified answering, additional retrieval, clarification and abstention.
A comparative evaluation protocol against LLM-only, standard RAG, hybrid RAG and generic adaptive RAG baselines.

III. RELATED WORK

A. Parkinson's Disease Literature and AI-Based Support

AI research in PD includes signal-based analysis of voice, gait, handwriting and imaging, as well as conversational systems intended to support access to disease information. These tasks differ from evidence-grounded literature question answering. A predictive model trained on a modality-specific dataset estimates an outcome from its input features; a literature QA system retrieves and synthesizes published evidence. Separate public datasets containing voice, gait or imaging records must not be combined as if they represented the same individuals unless participant-level linkage is explicitly documented.

Mukhtar et al. report a collaborative multi-agent conversational AI approach for clinical support in PD, combining generation, critique and synthesis with RAG over curated medical resources [4]. The present proposal therefore concentrates on evidence-context applicability and claim traceability rather than claiming the first PD RAG assistant.

B. Medical RAG, Adaptive Retrieval and Verification

Medical RAG systems retrieve biomedical or guideline material to ground generated answers [5], [6]. A retrieved passage can be relevant to a query yet insufficient to justify a clinical statement. Evaluation should distinguish retrieval relevance, factual support, citation correctness and clinical applicability. URGen [1] investigates uncertainty-guided retrieval and abstention in guideline-grounded medical question answering. MedRAGChecker [2] addresses claim-level verification for biomedical RAG, while medical abstention research [3] examines whether language models appropriately withhold answers under uncertainty. These are established methodological components; this proposal investigates how they interact with PD modality context and evidence applicability.

C. Research Gap and Positioning

The preliminary prior-art review identifies overlap in generic adaptive retrieval, biomedical claim verification, medical abstention and PD-oriented conversational RAG. It does not establish that no prior system combines these components with a structured PD-specific applicability model. The proposed research question is whether explicit representation of available, missing and unknown modalities, together with applicability assessment across population, stage, modality, outcome and design, provides a useful and auditable extension to existing medical RAG workflows. This remains a proposed contribution requiring systematic comparison and validation.

IV. DATASET AND PROBLEM SETTING

A. Task Definition

Given a natural-language question q, an optional structured context c and a corpus D of PD-related literature, the system produces an answer y, a set of atomic claims C, a set of source links E and a response action a. The action belongs to {answer with qualification, retrieve further, request clarification, abstain}. The system is designed for questions about published evidence and does not infer a diagnosis for an individual.

B. Context Representation

The context is represented as c = (p, s, M_A, M_M, M_U, o), where p denotes the stated population, s the stated disease stage, M_A the available modalities, M_M the explicitly missing modalities, M_U the unknown or unreported modalities, and o the outcome of interest. The three modality sets are kept distinct. An empty or unreported field is not automatically interpreted as a negative finding.

 This work is licensed under a Creative Commons Attribution 4.0 International License

TABLE I. CONTEXT REPRESENTATION AND MODALITY HANDLING

Field	Meaning	Example
Available (M_A)	Explicitly supplied in the question or vignette	Voice recording described as available
Missing (M_M)	Explicitly stated as not collected or unavailable	No gait recording available
Unknown (M_U)	Not specified in the supplied context	Disease stage not reported

C. Evidence Corpus and Benchmark Design

The proposed corpus will comprise eligible peer-reviewed PD studies, systematic reviews, consensus statements and authoritative clinical guidance. Candidate records may be identified through PubMed/MEDLINE and PubMed Central, with IEEE Xplore, Scopus and Web of Science included where institutional access permits. Records should retain DOI, PMID or PMCID where available, title, year, publication type, source, access status, population, disease stage, modality, outcome and study design. Search dates, exact queries, filters, duplicate counts and exclusion reasons must be recorded. The final corpus has not yet been collected or screened.

A proposed pilot benchmark contains 120 questions, subject to corpus coverage and reviewer availability: 60 answerable single-source questions, 30 multi-source synthesis questions, 15 questions involving conflicting or mixed evidence, and 15 questions for which context or evidence is insufficient. These are planned counts, not an existing dataset.

TABLE II. PROPOSED PILOT QUESTION CATEGORIES

Category	Count	Purpose
Single-source answerable	60	One suitable source
Multi-source synthesis	30	Synthesis across papers
Conflicting/mixed evidence	15	Explain disagreement
Insufficient context/evidence	15	Clarification or abstention

D. Data Integrity and Review

The benchmark is literature-derived and may use fictional, non-identifiable context vignettes to test modality handling. It will not contain real patient records or recordings. Review by a neurologist or suitably qualified clinician is proposed for question validity, clinical wording, answerability and safety. Independent annotation of a subset should be used to estimate agreement for claim support and applicability. Institutional ethics requirements must be confirmed with SRM Institute of Science and Technology before data collection.

V. PROPOSED METHODOLOGY

A. System Overview

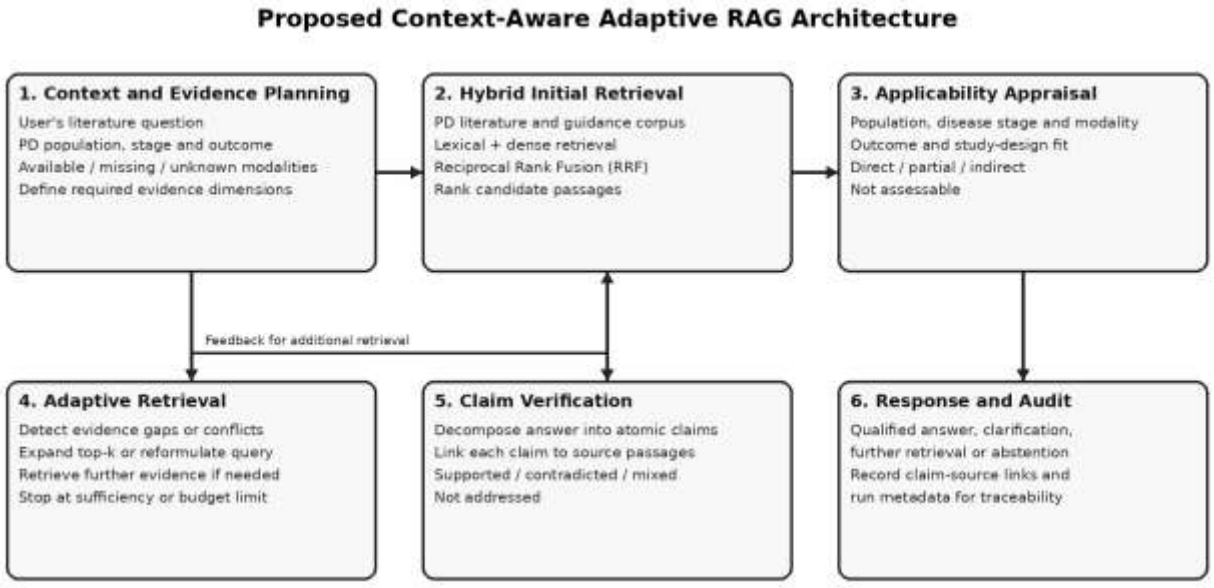


Fig. 1. Proposed architecture of the context-aware adaptive retrieval-augmented generation framework.

 This work is licensed under a Creative Commons Attribution 4.0 International License

The proposed workflow comprises eight stages: (1) context normalisation; (2) evidence requirement planning; (3) initial retrieval; (4) applicability appraisal; (5) adaptive retrieval; (6) claim decomposition and verification; (7) response-policy selection; and (8) audit-record generation. The framework is conceptual at this stage; no implemented system or measured output is claimed.

B. Context Normalisation and Initial Retrieval

The question is parsed into a structured representation containing research intent, population, PD stage, outcome and modality states. Each field is assigned an explicit value, missing marker or unknown marker. The parser must preserve the distinction between 'not available' and 'not mentioned'. Candidate evidence is retrieved from a frozen corpus using a hybrid strategy combining lexical retrieval and dense semantic retrieval. Reciprocal Rank Fusion (RRF) can combine rankings without requiring raw scores from different retrieval models to be directly comparable. Each chunk retains its document identifier and section or abstract location.

C. Evidence Applicability Assessment

Each candidate source is appraised on five dimensions: population fit, disease-stage fit, modality fit, outcome fit and study-design relevance. Each dimension may be labelled direct, partial, indirect or not assessable. These labels describe applicability to the question, not a universal ranking of study quality. Study limitations, sample characteristics and measurement protocol are retained as caveats.

TABLE III. EVIDENCE APPLICABILITY DIMENSIONS

Dimension	Question
Population	Does the study population match the question?
Disease stage	Is the reported stage relevant?
Modality	Does the study examine the relevant modality?
Outcome	Does the study measure the outcome asked about?
Study design	Can the design support the requested inference?

D. Adaptive Retrieval Controller

The controller examines whether the initial evidence set covers the required dimensions and whether sources disagree or leave material claims unresolved. If evidence coverage is insufficient and retrieval budget remains, it may expand top-k, reformulate the query or target a missing evidence dimension. If the question lacks essential context, the system should request clarification rather than retrieve indefinitely. If additional retrieval does not resolve the evidence gap, it should qualify or abstain.

A conceptual stopping rule is to stop when (i) each material evidence requirement has at least one applicable source, (ii) major conflicts are represented, or (iii) the retrieval budget is exhausted. Thresholds and stopping criteria must be operationalised and evaluated; they are not clinically validated thresholds.

E. Claim Verification and Response Policy

A draft response is decomposed into atomic factual claims. Each claim is linked to retrieved passages and assigned one of four support labels: supported, contradicted, mixed or not addressed. A citation is correct only when the cited passage substantively supports the attached claim. A source that merely discusses the same topic is not sufficient.

The response controller selects one of four actions: qualified answer when material claims have adequate support; further retrieval when evidence gaps may be addressable; clarification when essential context is absent or unknown; or abstention when the corpus does not support a sufficiently grounded response.

F. Audit Output

For reproducibility, each run should retain the question identifier, context fields, corpus snapshot, retrieved document and chunk identifiers, ranks, applicability labels, claim-to-source links, response action, model identifier, prompt version, timestamp, latency and API usage metadata where available. Sensitive patient information is outside the scope of this project.

VI. PROPOSED EVALUATION AND EXPECTED OUTCOMES

No implementation or experiments have been performed. This section defines the proposed comparison protocol and must not be interpreted as completed evaluation.

 This work is licensed under a Creative Commons Attribution 4.0 International License

TABLE IV. BASELINE AND COMPARISON SYSTEMS

ID	System	Description
S0	LLM-only	No external retrieval
S1	Standard RAG	Fixed top-k dense retrieval
S2	Hybrid RAG	BM25 and dense retrieval combined using RRF
S3	Generic adaptive RAG	Retrieval triggered by generic uncertainty or evidence gaps
S4	Proposed PD-context-aware RAG	Adds modality context, applicability appraisal, claim verification and response control

For a fair comparison, systems should use the same base language model, question set, corpus snapshot, generation settings and output budget where applicable. Both fixed-budget and adaptive-budget results should be reported. Thresholds and prompts should be tuned only on development data; the final test set must remain held out.

TABLE V. PROPOSED EVALUATION METRICS

Aspect	Metrics
Retrieval quality	Recall@5, Recall@10, Precision@k, MRR, nDCG
Claim grounding	Support, unsupported-claim and contradiction rates
Citation quality	Claim-to-source citation correctness
Applicability	Macro-F1 or agreement with expert labels
Answer quality	Expert-rated correctness, completeness and qualification
Abstention	Coverage, selective risk, abstention precision/recall
Efficiency	Retrieval rounds, API calls, tokens, latency and cost

A. Statistical Analysis and Ablations

The planned analysis uses paired question-level comparisons and bootstrap confidence intervals. Results should be stratified by question category and answerability. Coverage must be reported alongside selective risk. Planned ablations remove modality-state representation, applicability appraisal, adaptive retrieval and claim-level verification. No ablation result is currently available.

B. Expected Outcomes

The expected outcome is a formally specified, auditable workflow for context-aware evidence retrieval and literature question answering in PD. Evaluation is intended to determine whether the proposed context and applicability components improve evidence selection, claim support and appropriate abstention compared with the stated baselines. These are hypotheses to be tested, not established results.

VII. LIMITATIONS AND SAFETY

The principal limitation is that the proposed framework has not yet been implemented or empirically evaluated. Its effectiveness, computational cost and reproducibility remain unknown. The benchmark design and proposed question counts are targets rather than completed data collection. Expert review is planned but has not been obtained. The preliminary prior-art search is targeted rather than exhaustive; full-text comparison across relevant databases remains necessary before making a strong novelty claim.

Differences in cohort composition, disease stage, diagnostic criteria, acquisition protocols and outcome definitions may limit evidence transfer. Abstract-only records may omit methodological details required for applicability appraisal. Language-model services may change over time and introduce cost, latency, privacy and reproducibility constraints. Model outputs and applicability labels require human review.

The proposed system is intended only for literature navigation and evidence-grounded question answering. It is not a medical device, diagnostic system or substitute for a neurologist. It must not infer an individual's disease status, recommend treatment or combine records from unrelated datasets as though they belong to one person. Real patient data are excluded from the proposed benchmark.

VIII. CONCLUSION AND FUTURE WORK

This proposal presents a context-aware adaptive RAG framework for question answering over Parkinson's disease literature. The design makes modality availability explicit, separates topical relevance from evidence applicability, links atomic claims to source passages and provides a controlled choice among answering, further retrieval, clarification and abstention. Existing work already addresses generic medical adaptive retrieval, biomedical claim verification, medical abstention and Parkinson's conversational RAG; therefore, the proposed contribution is positioned narrowly around their structured integration with PD-specific context and applicability appraisal.

Future work will complete a systematic prior-art search, curate and clinically review the literature question benchmark, operationalise applicability labels and controller thresholds, implement the five comparison systems, and evaluate retrieval, claim grounding, applicability, abstention and efficiency using a held-out test set. Only after these steps can the framework's empirical value be assessed.

ACKNOWLEDGMENT

The authors acknowledge the guidance of Dr. Arulmozhi, Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India. The supervisor is acknowledged for academic guidance; authorship has not been assigned.

REFERENCES

[1] M. Hoi et al., "URGen: Uncertainty-guided retrieval-augmented generation for guideline-grounded medical question answering," in Findings of the Association for Computational Linguistics: EMNLP, 2026, doi: 10.18653/v1/2026.findings-emnlp.184.

[2] Y. Ji et al., "MedRAGChecker: Claim-level verification for biomedical retrieval-augmented generation," in Findings of the Association for Computational Linguistics: EMNLP, 2026. Verify publication metadata against the publisher record.

[3] A. Cocchieri et al., "LLMs (Almost) Never Abstain Under Medical Uncertainty," in Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics, 2026. Verify publication metadata.

[4] A. Mukhtar et al., "Collaborative multi-agent conversational artificial intelligence for clinical support in Parkinson disease," Parkinsonism & Related Disorders, 2026, doi: 10.1016/j.parkreldis.2026.108292.

[5] J. Dobreva et al., "RAGCare-QA: A benchmark dataset for evaluating retrieval-augmented generation pipelines in theoretical medical knowledge," Data in Brief, vol. 58, art. no. 112146, 2025, doi: 10.1016/j.dib.2025.112146.

[6] R. Kumar et al., "Healthcare retrieval-augmented generation evaluation: A scoping review," Journal of Medical Internet Research, 2026. Verify article number and DOI before submission.

 This work is licensed under a Creative Commons Attribution 4.0 International License