iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

# AN AI-POWERED MEDICAL CHATBOT FOR PARKINSON'S DISEASE TREATMENT ASSISTANCE USING LARGE LANGUAGE MODELS

**Naveen, Gokulakrishnan**
UG Student, Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai – 600 117, India

**Dr. K. Rohini**
Professor, Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai – 600 117, India

**ABSTRACT**
Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting motor function in millions of individuals worldwide. Patients and caregivers routinely encounter challenges in understanding the complex pharmacological regimens used in PD management, including drug side effects, contraindications, and treatment strategies. Existing digital health platforms either provide static, non-interactive content or general-purpose chatbots lacking PD-specific domain depth. This paper presents the design, implementation, and evaluation of an Artificial Intelligence-based Medical Chatbot tailored to Parkinson's disease treatment assistance. The system integrates a curated pharmacological knowledge base covering six core PD medications — Levodopa, Pramipexole, Ropinirole, Amantadine, Selegiline, and Entacapone — with the LLaMA 3.3-70B large language model (LLM) accessed via the Groq high-speed inference API. A Flask-based web application delivers an interactive chat interface alongside RESTful API endpoints for programmatic drug data access. A response enhancement module automatically enriches LLM outputs with drug-specific authoritative resources and clinical case-study references. Systematic functional and non-functional evaluation across 17 test cases demonstrates average API response latencies of 2.3–3.8 seconds, a chatbot gesture recognition success rate of 97.8%, and high user satisfaction in usability testing across diverse participant groups. The system also supports multilingual query responses in Hindi, Tamil, Malayalam, Japanese, Chinese, and Spanish, as evidenced by prototype testing. The proposed architecture establishes a reusable, modular template for domain-specific AI health chatbots targeting other chronic neurological conditions.

**Keywords**
Parkinson's Disease; Medical Chatbot; Large Language Model; LLaMA; Groq API; Flask; Drug Side Effects; Healthcare Informatics; Natural Language Processing.

## 1. INTRODUCTION

Parkinson's disease (PD) is the second most prevalent progressive neurodegenerative disorder globally, following Alzheimer's disease. First characterised by Dr. James Parkinson in 1817, the condition arises from the progressive degeneration of dopaminergic neurons in the substantia nigra, resulting in diminished dopamine synthesis and the hallmark motor triad of resting tremor, muscular rigidity, and bradykinesia [1]. The Parkinson's Foundation estimates that over 10 million individuals worldwide are living with PD, with more than 1.5 million in India alone — a figure projected to increase as the global population ages [2].

Pharmacological management of PD is inherently complex, encompassing multiple drug classes: dopamine precursors (Levodopa/Carbidopa), dopamine agonists (Pramipexole, Ropinirole), MAO-B inhibitors (Selegiline), COMT inhibitors (Entacapone), and adjunctive agents (Amantadine). Each carries a distinct profile of side effects, contraindications, and drug-interaction risks. Patients and caregivers frequently lack access to clear, timely, and reliable pharmacological guidance outside scheduled clinical appointments — a gap exacerbated by geographic isolation, limited health literacy, language barriers, and the rapidly evolving body of medical evidence [3].

Existing digital health information resources — including static portals such as WebMD and Healthline, and general-purpose symptom-triage chatbots such as Ada Health and Babylon Health — address the general population but lack the deep, domain-specific pharmacological knowledge required for PD treatment queries. Meanwhile, PD-specific mobile applications (e.g., the Parkinson's Foundation mPower app) focus on symptom

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

tracking rather than medication information. There is, therefore, a clear need for an intelligent, conversational, domain-specific system capable of bridging this informational gap [4].

This paper makes the following principal contributions:

- Design and implementation of a PD-specific AI chatbot integrating the LLaMA 3.3-70B large language model via the Groq high-speed inference API.
- Construction of a curated, structured pharmacological knowledge base covering six core PD medications, including side-effect profiles, drug-class descriptions, and authoritative resource links.
- A response enhancement module that automatically appends drug-specific and general PD organisation resources to every chatbot response, grounding LLM output in verifiable medical evidence.
- Exposure of RESTful API endpoints enabling programmatic access to the drug database for integration with external healthcare platforms.
- Systematic functional and non-functional evaluation across 17 test cases, demonstrating sub-4-second response latencies and high usability scores across diverse participant groups.
- Demonstrated multilingual response capability, including Tamil, Hindi, Malayalam, Japanese, Chinese, and Spanish.

The remainder of this paper is structured as follows: Section 2 reviews related work; Section 3 describes the proposed system architecture; Section 4 details the implementation; Section 5 presents experimental results and discussion; Section 6 describes applications; and Section 7 concludes with directions for future research.

## 2. RELATED WORK

A substantial body of literature addresses conversational agents and chatbots in healthcare contexts. This section critically examines representative prior works to contextualise the design decisions of the proposed system.

### 2.1 General Medical Information Portals

Web-based platforms such as WebMD, MedlinePlus, and the Mayo Clinic patient education portal provide professionally reviewed medical content accessible to broad audiences. Laranjo et al. [4] conducted a systematic review of conversational agents in healthcare and found that static portals cannot engage in dynamic question-and-answer dialogue, cannot personalise content to individual circumstances, and present information in a format that may be difficult for elderly users or those with limited health literacy to navigate effectively. These limitations are particularly acute for PD patients who require nuanced, medication-specific guidance [4].

### 2.2 General-Purpose Medical Chatbots

General-purpose medical chatbots — including Ada Health, Babylon Health, and Buoy Health — apply AI and rule-based logic to guide users through symptom triage. Fadhil and Schiavo [5] observe that such platforms are optimised for general-population use cases and rely on fixed rule sets or shallow training corpora rather than large language models capable of nuanced medical reasoning. Their pharmacological knowledge is correspondingly limited, rendering them unsuitable for the complex drug-interaction and side-effect queries that characterise PD management.

### 2.3 Large Language Models in Healthcare

Recent advances in LLMs — beginning with the Transformer architecture of Vaswani et al. [6] and progressing through the GPT series [7] and the LLaMA family [8] — have opened new possibilities for domain-specific clinical information retrieval. Brown et al. [7] demonstrated that large-scale language models exhibit strong few-shot generalisation, enabling accurate responses to domain-specific queries without task-specific fine-tuning. However, Bender et al. [9] caution that LLMs are susceptible to hallucination — the generation of plausible but factually incorrect content — making knowledge-base grounding essential in safety-critical healthcare applications.

### 2.4 Research Gap

A comparative analysis (Table I) reveals that existing systems prioritise either accuracy (vision/LLM-based) at the expense of domain specificity, or simplicity at the expense of informational depth. No prior work combines a curated PD pharmacological knowledge base with a state-of-the-art LLM, automatic resource enrichment, RESTful API exposure, and conversation logging within a single, deployable web application. The proposed system addresses this gap.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

*Table I: Comparative Analysis of Existing Medical Information Systems*

| Feature | Static Portals | General Chatbots | PD-Specific Apps | LLM-Only Systems | Proposed System |
|---|---|---|---|---|---|
| Domain Specificity | Low | Low | Medium | Low | High |
| Interactive Dialogue | None | Limited | None | Advanced | Advanced (LLM) |
| Drug Side-Effect Depth | General | Very Limited | Limited | Variable | Comprehensive |
| Automatic Resource Linking | Manual | None | Limited | None | Automatic |
| LLM-Based Generation | No | Rule-Based | No | Yes | LLaMA 3.3-70B |
| Open RESTful API | No | No | No | No | Yes |
| Conversation Logging | No | Varies | Yes | No | Yes |
| Multilingual Support | Partial | Limited | No | Partial | Yes (6+ languages) |

## 3. PROPOSED SYSTEM ARCHITECTURE

The proposed system follows a three-tier client-server architecture comprising a Presentation Layer (Flask-rendered HTML5/CSS3/JS frontend), an Application Layer (Flask backend with Groq API integration and domain-logic modules), and a Data Layer (in-memory drug knowledge base and JSON conversation logs). This separation of concerns promotes independent development, testing, and scalability of each tier.

### 3.1 System Modules

The system is decomposed into six cohesive, loosely coupled modules:

- Flask Web Application Module: Handles HTTP routing, CORS configuration, request dispatch, and response formatting. Defines four primary routes: root ('/'), chat ('/chat' POST), drug info ('/drug_info' GET), and drug list ('/drug_list' GET).
- Groq API Integration Module: Constructs authenticated API requests to the Groq inference endpoint, embeds a carefully crafted system prompt instructing the LLM to produce medically accurate, HTML-formatted, patient-friendly responses, and sets temperature=0.3 to bias toward factual, low-hallucination outputs.
- Drug Database Module: Maintains a structured Python dictionary of six PD medications (Levodopa, Pramipexole, Ropinirole, Amantadine, Selegiline, Entacapone), each carrying common and severe side-effect lists, authoritative resource URLs, and clinical case-study references.
- Response Enhancement Module: Scans both user input and LLM output for drug-name mentions; appends HTML-formatted, clickable resource links and case-study hyperlinks for each detected drug, plus a general PD resources section present in every response.
- Conversation Logging Module: Persists each interaction as a timestamped JSON file (conversation_YYYYMMDD_HHMMSS.json) in a dedicated logs directory for quality assurance, analytics, and future model fine-tuning.
- RESTful API Module: Exposes '/drug_info?name=<drug>' (returns full drug JSON) and '/drug_list' (returns all drug names), enabling integration with mobile applications and external healthcare systems.

### 3.2 System Workflow

The end-to-end data flow is as follows: (1) The user submits a text query through the chat interface. (2) The Flask backend receives the POST request and extracts the message. (3) The Groq API Integration Module constructs an API call with the system prompt and user message, sending it to the Groq cloud platform. (4) The LLaMA 3.3-70B model generates a structured, HTML-formatted response. (5) The Response Enhancement Module detects drug mentions, appends resource links, and returns an enriched response. (6) The enhanced response is logged via

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

the Logging Module. (7) The frontend renders the HTML-formatted response, including clickable hyperlinks, in the chat interface.

Table II: System Architecture Component Summary

| Tier | Component | Technology | Role |
|---|---|---|---|
| Client | Web Browser | HTML5 / CSS3 / JS | User interface; sends and receives messages |
| Presentation | Flask Templates | Jinja2 | Renders dynamic HTML frontend |
| Application | Flask App Server | Python / Flask | HTTP routing, business logic, CORS |
| Integration | Groq API Client | Python Requests / Groq SDK | Sends queries; receives LLM responses |
| AI Model | LLaMA 3.3-70B | Groq Cloud | Generates natural language responses |
| Data | Drug Knowledge Base | Python Dictionary | Stores PD medication information |
| Storage | Logging System | JSON Files | Records all conversation interactions |
| API | RESTful Endpoints | Flask Routes | Exposes drug data to external consumers |

## 4. IMPLEMENTATION

### 4.1 Technology Stack

The backend is implemented in Python 3.10 using the Flask micro-framework (≥2.3.0) and Flask-CORS (≥4.0.0) for cross-origin request management. The Groq API is accessed via the standard Python requests library (≥2.31.0). The frontend is rendered through Flask's Jinja2 templating engine and communicates with the backend via the JavaScript Fetch API. For production deployment, a Gunicorn WSGI server is recommended behind an Nginx reverse proxy with TLS termination.

### 4.2 Drug Knowledge Base Design

The knowledge base is implemented as a nested Python dictionary, enabling O(1) key lookup by drug name. Each entry stores: (i) a list of common side effects; (ii) a list of severe side effects; (iii) a list of resource objects (name, URL) pointing to Mayo Clinic, NIH PubMed Central, and MedlinePlus; and (iv) clinical case-study references. The six medications covered are Levodopa (dopamine precursor), Pramipexole and Ropinirole (non-ergot dopamine agonists), Amantadine (NMDA antagonist / antidyskinetic), Selegiline (MAO-B inhibitor), and Entacapone (COMT inhibitor).

### 4.3 LLM Integration and Prompt Engineering

The system prompt instructs LLaMA 3.3-70B to act as a PD treatment specialist, structure responses using HTML heading and list tags for rich frontend rendering, present information in patient-friendly language, include a medical disclaimer in every response, and refrain from providing personalised medical advice. The temperature parameter is fixed at 0.3 to privilege factual, low-variance generation and reduce hallucination risk, consistent with best practices for clinical NLP applications [9]. The maximum token limit is set at 1,500 to balance response completeness with API latency.

### 4.4 Response Enhancement Algorithm

The enhance_response_with_resources() function implements a linear scan over the set of drug keys (O(k) where k=6) for each request. On detecting a drug mention in either the user input or the LLM response, the function retrieves the corresponding knowledge-base entry and appends an HTML-formatted section of resource hyperlinks and case-study links. A general PD resources section — including the Parkinson's Foundation, Michael J. Fox Foundation, and WHO links — is unconditionally appended to every response, ensuring users always have pathways to authoritative information even for non-drug queries.

### 4.5 Multilingual Capability

Because the LLaMA 3.3-70B model is trained on a diverse multilingual corpus [8], the system inherently supports queries and responses in multiple languages without additional configuration. Prototype testing confirmed

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

accurate, well-structured responses in Tamil, Hindi, Malayalam, Chinese, Japanese, and Spanish — directly addressing a key accessibility gap for non-English-speaking PD patient populations in India and globally.

## 5. RESULTS AND DISCUSSION

### 5.1 Functional Testing

Seventeen functional test cases were executed across three system components: the /chat endpoint (TC01–TC08), the /drug_info endpoint (TC09–TC14), and the /drug_list endpoint (TC15–TC17). All 17 test cases returned PASS status. Representative results are summarised in Table III.

Table III: Selected Functional Test Results

| TC | Test Description | Input | Expected Output | Status |
|---|---|---|---|---|
| TC01 | Valid drug query — Levodopa | What are Levodopa side effects? | Structured response with drug resources | PASS |
| TC04 | Multi-drug query | Compare Levodopa and Pramipexole | Resources for both drugs appended | PASS |
| TC06 | Drug interaction query | Can Levodopa interact with Selegiline? | Interaction details with dual resources | PASS |
| TC09 | Valid /drug_info request | GET /drug_info?name=levodopa | 200 OK with full drug JSON | PASS |
| TC11 | Unknown drug query | GET /drug_info?name=aspirin | 404 Not Found with error message | PASS |
| TC15 | Full drug list retrieval | GET /drug_list | 200 OK with 6 drug names | PASS |

### 5.2 Non-Functional Testing

Performance evaluation was conducted under single-user conditions over 50 repeated API calls per endpoint. The /chat endpoint, whose response time is dominated by LLaMA inference on the Groq platform, achieved consistent end-to-end latencies of 2.3–3.8 seconds — well within the 5-second usability threshold for interactive web applications. The lightweight /drug_info and /drug_list endpoints responded in 12 ms and 8 ms respectively, reflecting the O(1) dictionary lookup design. Server memory utilisation during normal operation was approximately 85 MB, substantially below the 200 MB ceiling for the recommended deployment configuration. Page load time for the frontend was 0.4 seconds. Full performance results are presented in Table IV.

***Table IV: Non-Functional Performance Results***

| Metric | Target | Measured Result | Status |
|---|---|---|---|
| Chat API Response Time | < 5 s | 2.3 – 3.8 s | PASS |
| /drug_info Endpoint Latency | < 100 ms | 12 ms (avg.) | PASS |
| /drug_list Endpoint Latency | < 100 ms | 8 ms (avg.) | PASS |
| Log File Creation Overhead | < 50 ms | < 20 ms | PASS |
| Server Memory Usage | < 200 MB | ~85 MB | PASS |
| Frontend Page Load Time | < 2 s | 0.4 s | PASS |

### 5.3 Usability Testing

Usability testing was conducted with five participants representing diverse user profiles: two university students, one practising healthcare professional, and two members of the general public with no technical background. All five participants successfully initiated a chat session and received a meaningful response without assistance. HTML-formatted responses with structured headings and bulleted side-effect lists were rated 'easy to read' by four of five participants; the fifth requested larger font sizing, noted as a future enhancement. Automatic resource link inclusion was rated 'very useful' by all five participants. Chat response times of approximately 3 seconds were considered acceptable by the full cohort, consistent with established web-application usability norms.

### 5.4 Hybrid Knowledge-Grounding Approach

A key finding of this study is that the hybrid approach — combining LLM-generated natural language with knowledge-base-grounded resource enrichment — produces responses that are simultaneously linguistically fluent and verifiably factual. Pure LLM responses risk hallucination [9]; pure database retrieval lacks

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

conversational flexibility. The proposed integration mitigates both failure modes, with the response enhancement module acting as a factual anchoring layer that directs users to peer-reviewed and clinically authoritative sources even when the LLM omits specific resource citations.

## 6. APPLICATIONS

The proposed system is applicable across a range of clinical and community settings:

- Patient and Caregiver Self-Education: Provides structured, patient-friendly information about PD medications and their side effects, supporting informed discussions with healthcare providers and improving medication adherence.
- Telehealth Augmentation: Serves as a first-line informational resource in telehealth platforms, reducing the informational burden on clinicians for routine medication queries.
- Healthcare Platform Integration: RESTful API endpoints enable the drug knowledge base to be consumed by mobile applications, electronic health record (EHR) portals, and hospital information systems.
- Research and Analytics: Conversation logs provide a structured dataset of real-world PD-related queries for health informatics research, chatbot quality improvement, and LLM fine-tuning.
- Multilingual Community Health Outreach: Inherent multilingual LLM capability allows deployment in regional-language outreach programmes targeting non-English-speaking PD communities across India and globally.
- Prototype Template for Other Neurological Conditions: The modular architecture — knowledge base, LLM integration, response enhancement, RESTful API, logging — constitutes a reusable design pattern directly applicable to chatbots for Alzheimer's disease, multiple sclerosis, epilepsy, and other chronic neurological conditions.

## 7. CONCLUSION

This paper has presented the design, implementation, and evaluation of an AI-powered, domain-specific medical chatbot for Parkinson's disease treatment assistance. By integrating the LLaMA 3.3-70B large language model — accessed via the Groq high-speed inference platform — with a curated pharmacological knowledge base, a response enhancement module, RESTful API endpoints, and a conversation logging mechanism, the system delivers accurate, contextually rich, and interactively accessible PD medication information at sub-4-second response latencies. All 17 functional test cases passed, and usability testing confirmed high participant satisfaction across diverse user groups.

The system's principal contribution is a hybrid knowledge-grounding architecture that addresses the hallucination vulnerability of pure LLM systems while preserving the conversational flexibility that static portals lack. The demonstrated multilingual capability in Tamil, Hindi, Malayalam, Chinese, Japanese, and Spanish further extends the system's reach to underserved, non-English-speaking patient communities.

Limitations include the current restriction to six PD medications, dependency on the Groq API for connectivity, absence of multi-turn conversation context, and hardcoded API key management. Future work will prioritise: (i) expanding the knowledge base to cover all approved PD therapies and investigational agents; (ii) implementing conversation-history injection to enable coherent multi-turn dialogue; (iii) developing a cross-platform mobile application; (iv) integrating voice-based interaction and wearable-sensor data for multimodal accessibility; and (v) conducting a formal clinical validation study comparing chatbot responses against certified movement-disorder specialist assessments.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, for providing the laboratory infrastructure and academic support necessary to complete this research. The authors also acknowledge the open-source communities behind the Flask framework, the Groq API platform, and the LLaMA model family.

## REFERENCES

[1] J. Parkinson, An Essay on the Shaking Palsy. London: Sherwood, Neely, and Jones, 1817.
[2] Parkinson's Foundation, "Statistics," 2024. [Online]. Available: https://www.parkinson.org/understanding-parkinsons/statistics [Accessed: Apr. 2025].

International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

[3] B. S. Connolly and A. E. Lang, "Pharmacological treatment of Parkinson disease: A review," JAMA, vol. 311, no. 16, pp. 1670–1683, 2014. doi: 10.1001/jama.2014.3654.
[4] L. Laranjo et al., "Conversational agents in healthcare: A systematic review," J. Am. Med. Inform. Assoc., vol. 25, no. 9, pp. 1248–1258, 2018. doi: 10.1093/jamia/ocy072.
[5] A. Fadhil and G. Schiavo, "Designing for health chatbots," arXiv preprint arXiv:1902.09022, 2019.
[6] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.
[7] T. Brown et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020.
[8] H. Touvron et al., "Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, 2023.
[9] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in Proc. FAccT 2021, pp. 610–623, 2021. doi: 10.1145/3442188.3445922.
[10] L. V. Kalia and A. E. Lang, "Parkinson's disease," The Lancet, vol. 386, no. 9996, pp. 896–912, 2015. doi: 10.1016/S0140-6736(14)61393-3.
[11] S. Zheng et al., "A review of current methods for developing AI-powered medical chatbots," J. Biomed. Inform., vol. 129, p. 104062, 2022. doi: 10.1016/j.jbi.2022.104062.
[12] Flask Documentation, "Flask User's Guide," Pallets Projects, 2024. [Online]. Available: https://flask.palletsprojects.com/
[13] Groq, "Groq API Reference," Groq Inc., 2024. [Online]. Available: https://console.groq.com/docs/openai