

Article

Not peer-reviewed version

Electrochemistry Assistant: A Retrieval-Augmented Question Answering System for Electrochemical Information Retrieval

[Aravindan Natarajan](#)^{*} and Preethi Sankaranarayanan

Posted Date: 21 May 2025

doi: [10.20944/preprints202505.1674.v1](https://doi.org/10.20944/preprints202505.1674.v1)

Keywords: Retrieval Augmented Generation (RAG); Large Language Models (LLM); Gemini; ChromaDB; electrochemistry



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Electrochemistry Assistant: A Retrieval-Augmented Question Answering System for Electrochemical Information Retrieval

Aravindan Natarajan ^{1,*} and Preethi Sankaranarayanan ²

¹. Independent Researcher, Bangalore, Karnataka, India – 560109

². Independent Researcher, Madurai, Tamil Nadu, India – 625014

* Correspondence: anatarajank@gmail.com The correct email ID is: anatarajank@gmail.com

Abstract: This article details the development and evaluation of an Electrochemistry Assistant, a question-answering system engineered to furnish precise and comprehensive responses to inquiries about electrochemistry. The system implements a Retrieval Augmented Generation (RAG) methodology, capitalizing on the capabilities of Google's Gemini model and the ChromaDB vector database. It extracts pertinent information from a curated corpus of PDF documents, generates embedding vectors for both the documents and user queries, and retrieves relevant documents based on semantic similarity. Subsequently, the Gemini model utilizes the retrieved documents as contextual information to generate informative and user-centric answers. The performance of the system was evaluated using a dual approach: employing accuracy, precision, recall, and F1 score metrics on a defined set of test questions, and through a manual review by domain experts to assess the accuracy and relevance of the generated responses. The findings highlight the potential of RAG in constructing specialized question-answering systems for scientific domains such as electrochemistry, particularly in the retrieval and synthesis of information from technical documentation. The complete source code is freely available and accessible via the following repository: <https://github.com/anatarajank/Electrochemistry-Assistant>.

Keywords: Retrieval Augmented Generation (RAG); Large Language Models (LLM); Gemini; ChromaDB; Electrochemistry

1. Introduction

Electrochemistry, the study of chemical transformations involving electron transfer, plays a pivotal role across diverse scientific, industrial, and technological disciplines [1,2]. Researchers, students, and professionals frequently require access to and processing of extensive and complex information on electrochemical methodologies. Nevertheless, conventional information retrieval systems, often predicated on keyword-based searches, may not adequately capture the semantic subtleties inherent in user queries and may fail to yield extensive responses.

In recent years, Retrieval Augmented Generation (RAG) has garnered considerable attention as a robust methodology for constructing advanced question-answering systems [3,4]. This paradigm synergistically integrates the proficiencies of large language models (LLMs) in natural language comprehension and generation with the capacity to retrieve pertinent information from external knowledge repositories. RAG systems characteristically involve generating embedding vectors for both documents and queries, storing them within a vector database, and retrieving relevant documents based on semantic similarity. The retrieved documents then serve as contextual information for the LLM to generate responses. The effectiveness of RAG has been substantiated across various domains, including open-domain question answering [5], scientific literature analysis

[6–8], and customer support [9], demonstrating notable enhancements in accuracy and comprehensiveness compared with traditional keyword-based methodologies.

Chemistry-focused question-answering systems have explored knowledge graphs, as seen in the development of Marie [10–15]. Early versions, including initial iterations of Marie, were largely template-based [15,16]; newer implementations utilize pre-trained language models for question answering. More recently, Large Language Model (LLM) agents that use Retrieval-Augmented Generation (RAG) show promise for enhancing chemical information retrieval. For example, Chen et al. [17] developed Chemist-X, a RAG-based agent for reaction condition recommendations. Within the specialized field of electrochemistry, a self-hosted RAG application attributed to Robert Chukwu [18], which utilizes Go, templ, OpenAI, and Pinecone, provides an example of responding to user queries concerning electrochemical impedance spectroscopy. However, thorough documentation specifying the application's precise capabilities and limitations is currently unavailable.

This paper describes the development of an Electrochemistry Assistant, a RAG-based question-answering system specifically conceived for electrochemical information retrieval, focusing on prevalent electroanalytical techniques. The system harnesses the Gemini model to generate both document and query embeddings and employs ChromaDB for the efficient storage and retrieval of these embeddings. The system's performance is evaluated on a defined set of test questions employing accuracy, precision, recall, and F1 score metrics, and its efficacy in providing accurate and contextually applicable responses was demonstrated. Furthermore, the accuracy and contextual applicability of the generated responses were corroborated through a manual review by domain experts and automated semantic similarity analysis.

2. Methodology

The operational flow of the system, from electrochemical inquiry to response generation, is depicted in the flowchart in Fig. 1.

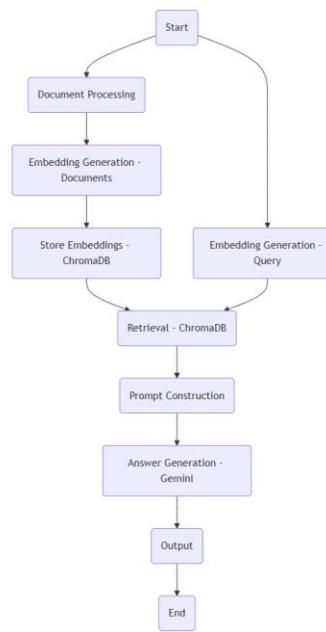


Figure 1. Information flow within the Electrochemistry Assistant, from query input to response generation.

This visual representation complements the textual description of our methodology.

2.1. Implementation Details

The Electrochemistry Assistant was implemented using Python. This paper presents an expanded version of a system initially developed by the first author within a Kaggle environment [19]. The Kaggle notebook environment facilitated efficient coding, access to computational

resources, and collaboration during the development process. Successful execution of the system requires users to create a Gemini API key, following the procedure outlined elsewhere [20]. The source code is available at both as a Kaggle notebook and in the project's Github repository [21].

2.2. Data Collection and Preprocessing

The Electrochemistry Assistant employs a curated corpus of PDF documents about electrochemistry techniques as its knowledge repository. These documents were acquired from a reputable online resource, Analytical Electrochemistry: The Basic Concepts [22]. The documents are stored in the ChromaDB database as shown in the Table.1 below:

Table 1. Corpus of the reference PDF documents and their storage information in the ChromaDB.

Document ID	Document Index	Document Name	Technical Notes On
Document1	0	PDF-16-WorkingElec.pdf	Different types of working electrodes and their preparation
Document2	1	PDF-10-CV-ImpParam.pdf	Important Parameters in Cyclic Voltammetry
Document3	2	PDF-6-Chronoamp.pdf	Chronoamperometry
Document4	3	PDF-9-CVintro.pdf	Introduction to Cyclic Voltammetry
Document5	4	PDF-12-CV-Additional.pdf	Non-Faradaic processes and their measurements
Document6	5	PDF-11-CV-Coupled.pdf	Chemical reactions coupled to electron transfer.
Document7	6	PDF-8-LSV.pdf	Linear Sweep Voltammetry
Document8	7	PDF-7-Chronocoul.pdf	Chronocoulometry
Document9	8	PDF-27-CareandFeeding.pdf	Care and Feeding of Electrodes
Document10	9	PDF-13-ASV.pdf	Anodic Stripping Voltammetry

The pdfminer.six library was utilized to extract textual data from the PDF files, which were subsequently preprocessed to eliminate extraneous content and formatting.

2.3. Embedding Generation

Embedding vectors are generated for both the extracted textual data from the documents and user queries through the Gemini API. The Gemini model, specifically trained for embedding generation, captures the semantic content of text and represents it as a vector within a multi-dimensional vector space. The embedding vectors for the documents are stored in a ChromaDB vector database, thereby enabling efficient search and retrieval based on semantic similarity.

2.4. Retrieval Process



Upon the formulation of a user query, the system initially generates an embedding vector for the query via the Gemini API. This query embedding vector is then employed to query the ChromaDB database for documents exhibiting similar embedding vectors. The most pertinent documents, as determined by cosine similarity metrics, are retrieved and furnished as contextual information for the response generation process.

2.5. Answer Generation

The Gemini model generates the final response to the user's query. The retrieved documents are integrated into a meticulously structured prompt, which is then input into the model. The prompt serves to guide the model, instructing it to synthesize an inclusive and user-centric answer grounded in the retrieved information. The model's sophisticated language comprehension and generation capabilities enable it to synthesize information from the retrieved documents and produce a coherent and informative response.

3. Evaluation

This section evaluates the Electrochemistry Assistant's performance in addressing user inquiries about electrochemistry techniques. The assessment focuses on the system's capacity to retrieve pertinent documents and generate accurate and informative responses.

3.1. Evaluation Dataset

A curated list of test inquiries shown in the table ST1 of the Supplementary Information and stored as list variables called queries (labelled Q0-Q9) and complex_queries (labelled CQ0-CQ9) in the Kaggle Notebook was compiled to evaluate the performance of the electrochemistry assistant. They target diverse electrochemical concepts and methodologies discussed within the source documentation. The inquiries were selected to assess the system's capacity to process heterogeneous query types and retrieve relevant information from the document corpus. For each inquiry, a specific set of documents was designated as "relevant" based on their content. This dataset served as the established ground truth for evaluating retrieval efficacy.

Furthermore, the simple queries are designed to evaluate the system's ability to generate accurate definitions based on the provided documents. Conversely, the complex queries are constructed to assess the system's ability to identify context across multiple documents and synthesize the requisite information. It is worth noting that the queries CQ7, CQ8, and CQ9 were specifically included to test the system's limitations, as the corpus does not contain documents directly relevant to these queries.

3.2. Evaluation Metrics

The performance of the Electrochemistry Assistant has been evaluated employing the following metrics:

- I. **Retrieval Accuracy:** This metric quantifies the system's proficiency in accurately retrieving relevant documents for each inquiry. It is computed as the ratio of relevant documents retrieved to the total number of relevant documents for that inquiry. Higher accuracy values indicate superior retrieval performance.
- II. **Precision:** Precision quantifies the proportion of correctly retrieved documents among the entirety of retrieved documents. It reflects the system's capacity to avoid the retrieval of extraneous documents.
- III. **Recall:** Recall quantifies the proportion of correctly retrieved documents among the entirety of relevant documents. It indicates the system's capacity to identify all relevant documents.

IV. **F1-score:** This metric integrates precision and recall, providing a more holistic evaluation of the system's performance. It represents the harmonic mean of precision and recall. A higher F1-score signifies enhanced overall retrieval performance.

V. **Mean Reciprocal Rank (MRR):** This metric evaluates the rank of the first relevant document retrieved for each query. The reciprocal rank of the first relevant document is calculated for each query, and the MRR is the average of these reciprocal ranks across all queries. A higher MRR indicates that relevant documents are ranked higher in the retrieval results.

VI. **Precision at K (P@K):** This metric evaluates the early retrieval performance by measuring the proportion of relevant documents present within the top K retrieved results. In this study, we report Precision at K=2 (P@2), K=3 (P@3), and K=5 (P@5) to provide a granular view of the system's ability to prioritize relevant information at different ranks. Specifically, P@2 assesses whether at least one of the top two retrieved documents is relevant, P@3 examines the relevance within the top three, and P@5 within the top five. Higher P@K values signify a greater concentration of relevant documents at the beginning of the retrieval list, indicating the system's effectiveness in quickly surfacing pertinent information to the user.

VII. **Normalized Discounted Cumulative Gain at K (NDCG@K):** This metric evaluates the quality of the document ranking by considering both the relevance of each retrieved document and its position within the ranked list, with greater weight assigned to relevant documents appearing earlier. In this study, we report the Normalized Discounted Cumulative Gain at K=2, 3, and 5 (NDCG@2, NDCG@3, and NDCG@5) to provide insight into the system's ability to rank meaningful information within the initial results effectively. Specifically, NDCG@2 assesses the ranking effectiveness within the top two retrieved documents, prioritizing instances where relevant documents are positioned at the highest ranks. Similarly, NDCG@3 examines the ranking quality within the top three results, again favoring the early appearance of relevant documents. Finally, NDCG@5 evaluates the ranking performance within the top five retrieved documents, emphasizing the significance of relevant documents in the initial positions. Higher NDCG@K values signify superior ranking performance, indicating a greater tendency for the most relevant documents to be positioned at the beginning of the retrieval list.

3.3. Baseline Retrieval Methods: BM25 and TF-IDF

In order to establish a comparative baseline for assessing the retrieval performance of the Electrochemistry Assistant, which employs semantic similarity within the ChromaDB vector database, two well-established traditional information retrieval methods, BM25 and TF-IDF, were also evaluated. These methods rely on term-based relevance scoring rather than semantic embeddings, offering a contrast to the approach employed by our primary system.

3.3.1. BM25 (Best Match 25)

BM25 is a ranking function used by search engines to estimate the relevance of a set of documents to a given search query. It is a bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document, regardless of the terms proximity within the document. BM25 considers the term frequency (TF) in the document, the inverse document frequency (IDF) of the term across the corpus, and the document length to calculate a relevance score.

We implemented BM25 using a standard library rank_bm25, applying common parameter settings to optimize its performance on our electrochemical document corpus.

3.3.2. TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF is another fundamental technique in information retrieval that assigns a weight to each term in a document based on its term frequency in that document and its inverse document frequency across the entire collection of documents. Terms that appear frequently in a specific document but infrequently in the rest of the corpus are assigned higher weights, indicating their importance to that document's content. We implemented TF-IDF by calculating the TF-IDF scores for all terms in our document corpus and then using these scores to rank documents based on the presence and weight of the query terms within them.

The performance of the Electrochemistry Assistant, leveraging ChromaDB, will be directly compared against the retrieval performance of these BM25 and TF-IDF baseline methods using the same evaluation dataset and metrics outlined in Section 3.1. This comparison will highlight the relative strengths and weaknesses of semantic versus term-based retrieval approaches in the context of electrochemical information retrieval.

3.4. Semi-Quantitative Assessment of the Generated Answers

We undertook automated and manual assessments to comprehensively evaluate the generated responses. The automated evaluation employed a custom calculate_f1_score() function to assess the semantic similarity between the generated answer and the retrieved passages. The calculate_f1_score() function employs a pre-trained sentence transformer model ('all-mpnet-base-v2') to generate sentence embeddings for both the generated answer and the retrieved passages, thus facilitating semantic similarity assessment. Subsequently, it computes the cosine similarity between the answer embedding and each passage embedding to identify suitable passages based on a predefined relevance threshold (0.70). Utilizing the sets of significant and retrieved passages, the function calculates precision and recall, ultimately resulting in the computation of the F1-score—a balanced metric that reflects the system's ability to generate answers semantically aligned with the retrieved context. This evaluation methodology leverages the inherent semantic understanding of the pre-trained model to provide an objective measure of answer quality.

Complementing this quantitative analysis, the generated responses underwent a manual review by the authors, both of whom possess over a decade of expertise in electrochemistry. This semi-quantitative assessment focused on the following criteria, using a 5-point ordinal scale (1: Poor, 2: Fair, 3: Good, 4: Very Good, 5: Excellent) for each judgment:

I. Accuracy:

- **Correctness:** Is the generated answer factually correct and aligned with the information provided in the relevant source documents?
- **Completeness:** Does the answer provide all the necessary information to fully address the user's question?
- **Precision:** Is the answer concise and focused on the specific question asked, or does it contain irrelevant or redundant information?

II. Relevance:

- **Question Alignment:** Does the answer directly address the user's question, or does it deviate from the intended topic?

- **Information Need:** Does the answer satisfy the user's information need, providing the type of information they were seeking?
- **Contextual Appropriateness:** Is the answer appropriate for the context of the question and the user's likely intent?

III. Fluency and Coherence:

- **Grammaticality:** Is the answer grammatically correct and free of errors in sentence structure and word usage?
- **Clarity:** Is the answer easy to understand and follow, or is it confusing or ambiguous?
- **Logical Flow:** Does the answer present information logically and coherently, or is it disjointed or difficult to follow?

IV. Overall Quality:

- **Helpfulness:** How helpful is the answer in providing the user with the information they need?
- **Satisfaction:** How satisfied would a user be with the answer provided?
- **Trustworthiness:** Does the answer appear to be trustworthy and reliable?

The insights derived from this expert evaluation, based on these detailed criteria and the ordinal scale, provide a more nuanced and comprehensive understanding of the system's efficacy in generating high-quality and contextually appropriate answers.

4. Results

This section details the results of our evaluation of the Electrochemistry Assistant, specifically its ability to find applicable information and provide correct answers to user questions about electrochemistry techniques.

4.1. Retrieval Performance Evaluation

The system's capacity for relevant document retrieval was evaluated through the application of a predefined set of test queries, as detailed in Section 3.1. The evaluation incorporated a range of metrics, including accuracy, precision, recall, and F1-score, which are shown in the table ST2 in the Supplementary Information. On the other hand, ranking-sensitive measures such as Mean Reciprocal Rank (MRR), Precision at K (P@K), and Normalized Discounted Cumulative Gain at K (NDCG@K) are presented in Table 2.

4.1.1. Retrieval Performance

The evaluation of three distinct retrieval methods—ChromaDB, TF-IDF, and BM25—across a diverse set of queries reveals nuanced performance characteristics for each approach. In several instances (Q1, Q2, Q5, Q6, Q8, Q9, and CQ3), all three methods demonstrated perfect retrieval efficacy, achieving optimal scores across all evaluated metrics: Accuracy, Precision, Recall, and F1-Score.

ChromaDB exhibited a generally robust performance profile, consistently achieving high scores and demonstrating superior or comparable results to the other methods across a significant portion of the query set. TF-IDF also demonstrated commendable performance on numerous queries, frequently aligning with or closely approximating the results obtained by ChromaDB. However,

notable deficiencies were observed for TF-IDF on specific queries (Q4, CQ8, and CQ9), where performance declined substantially.

BM25 presented a more variable performance landscape. While achieving perfect retrieval on several queries (Q1, Q2, Q5, Q6, Q7, Q8, Q9, and CQ3), it also exhibited marked underperformance on others (Q3, Q4, CQ8, and CQ9), suggesting a sensitivity to specific query characteristics or document distributions.

Overall, ChromaDB appears to offer a more consistently high level of retrieval performance across the evaluated query set. While TF-IDF proves competitive in many scenarios, its susceptibility to performance degradation on certain queries warrants consideration. BM25, despite achieving optimal results on a subset of queries, demonstrates a less stable performance profile compared to ChromaDB and TF-IDF. Further investigation into the specific characteristics of the queries where performance disparities are most pronounced may yield valuable insights for optimizing retrieval strategies for each method.

4.1.2. Retrieval Ranking Performance

The efficacy of ranking of the three distinct document retrieval namely, ChromaDB, TF-IDF, and BM25 was evaluated using a suite of established metrics: Mean Reciprocal Rank (MRR), Precision at K=2 (P@2), Precision at K=3 (P@3), Precision at K=5 (P@5), Normalized Discounted Cumulative Gain at K=2 (NDCG@2), NDCG at K=3 (NDCG@3), and NDCG at K=5 (NDCG@5). These metrics are shown in Table 2 and collectively assess the ability of each method to retrieve relevant documents and, critically, to rank them effectively within the initial segments of the result lists.

Table 2. Retrieval Ranking Performance Metrics.

Evaluation Metrics	Chromadb	TF-IDF	BM25
MRR	0.94	0.93	0.81
P@2	0.67	0.61	0.5
P@3	0.48	0.48	0.41
P@5	0.31	0.29	0.29
NDCG@2	0.92	0.89	0.74
NDCG@3	0.94	0.94	0.77
NDCG@5	0.96	0.93	0.88

Mean Reciprocal Rank (MRR) evaluates the average rank of the first relevant document retrieved across all queries. A higher MRR signifies superior performance in surfacing relevant information quickly. Among the methods, ChromaDB achieved the highest MRR (0.94), indicating its greater effectiveness in positioning relevant documents at higher ranks. TF-IDF follows with an MRR of 0.93, while BM25 exhibits the lowest MRR (0.81).

Precision at K (P@K) measures the proportion of relevant documents within the top K retrieved results. At K=3, ChromaDB demonstrates the highest precision (0.67), followed by TF-IDF (0.61) and BM25 (0.50). At K=5, both ChromaDB and TF-IDF achieve a precision of 0.48, outperforming BM25 (0.41).

Normalized Discounted Cumulative Gain at K (NDCG@K) assesses the ranking quality by considering the relevance of each document and its position, with higher relevance and earlier positions contributing more to the score. ChromaDB exhibits the highest NDCG at K=2 (0.92) and

K=5 (0.96), and a high NDCG at K=3 (0.94), matching TF-IDF's performance at this level. TF-IDF shows a competitive NDCG at K=2 (0.89) and K=5 (0.93). BM25 consistently yields the lowest NDCG scores across all cutoff points (0.74 at K=2, 0.77 at K=3, and 0.88 at K=5).

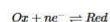
The evaluation of document retrieval ranking performance reveals that ChromaDB demonstrates the most consistently effective results across the majority of the employed metrics, indicating its superior ability to retrieve and rank relevant documents highly. TF-IDF provides a strong and often comparable performance, while BM25 generally exhibits a less effective ranking capability based on these evaluation scores. The consistently high NDCG values for ChromaDB and TF-IDF underscore their proficiency in prioritizing relevant information within the initial search results, a crucial aspect of effective document retrieval.

4.2. Response Generation Quality Assessment

A detailed evaluation of the generated responses was conducted through a dual assessment methodology, comprising both automated metrics and expert review. Illustrative response samples generated by the model are presented below: Fig. 2 displays the response to the simple query, "What is Cyclic Voltammetry?", and Fig. 3 displays the response to the complex query, "Compare and contrast linear sweep voltammetry and cyclic voltammetry, emphasizing the information that can be obtained from each technique and their respective advantages and limitations."

Cyclic voltammetry (CV) is an electrochemical technique where the potential applied to a working electrode is scanned linearly versus time, similar to linear sweep voltammetry (LSV). However, in CV, the potential scan is reversed at a specific point, called the switching potential, and scanned back to the initial potential, typically using a triangular waveform. This cycle allows for the observation of both the reduction and oxidation processes of an electroactive species.

In a CV experiment, the reduction of an oxidized species (Ox) at the electrode can be represented as:



where n is the number of electrons transferred and Red is the reduced form. The double arrow indicates that the reaction can proceed in both directions if the product is chemically reversible, meaning it's stable on the timescale of the experiment.

Important parameters in CV include:

- **Peak Locations:** The forward ($E_{f,p}$) and reverse ($E_{r,p}$) peak potentials, which can be used to calculate the formal potential (E^0):

$$E^0 = \frac{E_{f,p} + E_{r,p}}{2}$$

The peak separation, $\Delta E_p = |E_{f,p} - E_{r,p}|$, indicates electrochemical reversibility, with $\Delta E_p \approx \frac{0.058}{n} V$ for a reversible process.

- **Current Ratios:** The ratio of reverse to forward peak currents ($i_{p,r}/i_{p,f}$) is approximately 1 for a reversible process.

- **Scan Rate Dependence:** For reversible electron transfer, peak currents (i_p) increase linearly with the square root of the scan rate ($\nu^{1/2}$), as described by the Randles-Sevcik equation.

$$i_p = (2.69 \times 10^5) n^{3/2} A D^{1/2} C_i \nu^{1/2}$$

where A is the electrode area, D is the diffusion coefficient, and C_i is the concentration of the electroactive species.

It's also important to consider capacitive current, which arises from the rearrangement of ions in the double layer, and solution resistance, which causes a potential drop (Ohmic drop) described by $E_{IR} = i_{cell} R_s$, where i_{cell} is the cell current and R_s is the solution resistance. High scan rates or low dielectric solvents can exacerbate the effects of solution resistance.

CV is also useful for studying chemical reactions coupled to electron transfer, such as $E_r C_i$, $E_r C_r$, and $E_r C'_i$ mechanisms, where E represents an electron transfer step and C represents a chemical step. The changes in peak currents and potentials as a function of scan rate can provide insights into the kinetics and mechanisms of these reactions.

Figure 2. A sample response generated for the query "What is cyclic voltammetry?".

As evidenced in Fig. 2, the model generated a well-structured response to the aforementioned query, drawing upon information within the document corpus. Notably, the response to the complex query, illustrated in Fig. 3, exhibited greater comprehensiveness, likely due to a larger number of pertinent documents available to the model. However, minor inline LaTeX formatting inconsistencies were observed for some equations within this response and others. For the reader's detailed examination, the sample responses generated for all test queries are available in the project's GitHub repository [21].

Linear Sweep Voltammetry (LSV) and Cyclic Voltammetry (CV) are both potentiodynamic electrochemical techniques where the potential applied to a working electrode is varied linearly with time, and the resulting current is measured. However, they differ in the potential waveform applied and the information that can be extracted.

Linear Sweep Voltammetry (LSV):

In LSV, the potential is scanned in one direction only, from an initial potential to a final potential. The resulting current-potential profile is peak-shaped for a redox-active species undergoing electron transfer.

- Information Obtained:** LSV can be used to determine the formal potential (E°), number of electrons transferred (n), diffusion coefficient (D), and concentration (C) of the electroactive species for a reversible system. A plot of peak current (i_p) versus the square root of the scan rate ($v^{1/2}$) is linear for a diffusion-controlled redox process and serves as a diagnostic tool. The peak potential (E_p) for reversible electron transfer does not change with the scan rate. The separation between $E_{p/2}$ (potential at half peak height) and E_p can also indicate reversible electron transfer, with $|E_p - E_{p/2}| = \frac{RT}{4n} \text{ mV}$.
- Advantages:** LSV is relatively simple to perform and interpret for systems with straightforward redox behavior.
- Limitations:** LSV provides limited information about the stability of the products formed during the electron transfer reaction. It is less useful for studying complex reaction mechanisms.

Cyclic Voltammetry (CV):

In CV, the potential is scanned from an initial value to a switching potential, and then the scan is reversed back to the initial potential, typically using a triangular waveform.

- Information Obtained:** CV provides information about the redox potentials of the electroactive species, the reversibility of the electron transfer reaction, and the stability of the electrochemically generated species. The formal potential (E°) can be estimated by averaging the forward and reverse peak potentials: $E^\circ = \frac{1}{2}(E_{p,forward} + E_{p,reverse})$. The electrochemical reversibility can be determined from the peak separation (ΔE_p), with $\Delta E_p = \frac{RT}{4n} \text{ V}$ for a reversible process. The ratio of the reverse to forward peak currents ($i_{p,reverse}/i_{p,forward}$) is approximately 1 for a reversible process where the product is stable. The dependence of peak currents on the scan rate (i_p vs $v^{1/2}$) can provide insights into the kinetics and mechanism of the electron transfer reaction.
- Advantages:** CV is excellent for qualitative analysis and mechanistic studies, allowing for the assessment of the reversibility and stability of redox processes.
- Limitations:** CV can be more complex to interpret than LSV, especially for systems with multiple redox couples or irreversible reactions.

Comparison:

Feature	Linear Sweep Voltammetry (LSV)	Cyclic Voltammetry (CV)
Potential Waveform	Linear scan in one direction	Triangular scan (forward and reverse)
Reversibility	Limited information	Assesses electrochemical reversibility
Product Stability	Limited information	Assesses product stability
Mechanistic Information	Less detailed	More detailed
Complexity	Simpler	More complex

Both techniques are affected by factors such as capacitive current and solution resistance, which can distort the voltammograms. Capacitive current can be minimized by background subtraction or by using techniques with different waveforms, such as differential pulse voltammetry (DPV). Uncompensated solution resistance (R_s) leads to an Ohmic drop ($E_{IR} = i_{cell}R_s$), which can be minimized by proper cell design (e.g., using a Luggin capillary) or compensation techniques. Large currents, high scan rates, or low dielectric solvents exacerbate the effects of solution resistance, leading to drawn-out curves and increased peak separations in CV.

Figure 3. A sample response generated for the query “Compare and contrast linear sweep voltammetry and cyclic voltammetry, emphasizing the information that can be obtained from each technique and their respective advantages and limitations.”.

The automated evaluation leveraged the F1-score, calculated based on the semantic similarity between the generated response and the retrieved source passages, as elaborated upon in Section 3.4. The results of this quantitative analysis are presented in Table 3. The elevated scores in most cases suggest a high degree of semantic overlap, indicating that the generated response effectively integrated and synthesized information from the significant source documents.

Table 3. Semantic Similarity F1-Scores.

Query ID	F1-Score	Query ID	F1-Score
Q0	0.89	CQ0	0.33
Q1	0.57	CQ1	0.57
Q2	0.57	CQ2	0.57
Q3	0	CQ3	0.33
Q4	0.57	CQ4	0.89
Q5	0.33	CQ5	0.89
Q6	0.57	CQ6	0.89
Q7	0	CQ7	0
Q8	0	CQ8	0
Q9	0.33	CQ9	0.33

Complementing this quantitative analysis, a detailed manual expert assessment of all generated responses was also carried out, and the results are summarized in Table ST3 in the Supplementary Information and Table 4. This qualitative assessment focused on critical dimensions of response quality, including linguistic fluency and coherence, factual accuracy and alignment with the retrieved

documents, and the overall contextual appropriateness concerning the user's initial query. The insights from this extensive expert evaluation offer valuable triangulation of the system's effectiveness in producing high-quality and contextually relevant answers across the entire query set. Moreover, an overall average score of 4 and above, as presented in Table 4, across all evaluation criteria, indicated that the Electrochemistry Agent performed as expected.

Table 4. Overall average score for the Electrochemistry Agent across all the manual evaluation criteria.

Criteria	Sub-Criteria	Average Score
Accuracy	Correctness	4.15
	Completeness	4.38
	Precision	4.38
Relevance	Question Alignment	4.3
	Information Need	4.28
	Contextual Appropriateness	4.3
Fluency and Coherence	Grammaticality	4.48
	Clarity	4.53
	Logical Flow	4.53
Overall Quality	Helpfulness	4.35
	Satisfaction	4.25
	Trustworthiness	4.43

Despite the system's generally high scores in the manual evaluation, the authors deem it pertinent to draw the reader's attention to the following observations.

1. Given the stochastic nature of the Gemini family of models, the generated response may exhibit variability across different iterations of the same query. While this was mitigated by setting the temperature parameter to 0.5, which facilitated the generation of relatively consistent responses for each query, readers should be aware that the response quality may vary upon re-execution.
2. For simple queries, abbreviations such as Ox and Red, once defined within a response, were not consistently expanded or re-explained in subsequent responses to different queries executed sequentially.
3. In certain complex questions about electrode cleaning, the generated responses bore similarities to those produced for simpler queries.

4. In some instances, the provided answers and explanations did not entirely align with expected responses. For example, concerning the complex query CQ1, the system was unable to derive the relevant equation, likely due to the absence of a directly pertinent document within the corpus.
5. The model tended to generate responses, potentially representing instances of hallucination, for the complex queries CQ7, CQ8, and CQ9, for which no directly relevant passages or documents were manually identified within the document corpus. Nevertheless, the quality of these generated responses was deemed acceptable.

5. Discussion

The outcomes of our evaluation substantiate the Electrochemistry Assistant's capacity to furnish relatively accurate and exhaustive responses to a broad spectrum of electrochemical inquiries. The RAG methodology, capitalizing on the capabilities of the Gemini model and the ChromaDB database, enables the system to effectively retrieve appropriate information from a curated corpus of PDF documents and generate informative, contextually meaningful, and user-centric responses. The responses are also presented in a lucid and comprehensible format, thereby enhancing their accessibility for both technical and non-technical users. This capacity to provide contextually important and user-centric responses is crucial for the system's practical application in diverse settings.

While the system demonstrates efficacy on our curated dataset, its high accuracy and ranking metrics are likely atypical for broader information retrieval applications and warrant careful consideration. This observed performance may be attributed to several factors: the focused scope of our document corpus (fundamental electrochemistry), the direct nature of our test inquiries (seeking explicit definitions and explanations), and the constrained size of our initial knowledge base (ten PDF documents). Furthermore, the limited breadth of electrochemical topics due to the small repository and the dependence of response accuracy on the source document quality are inherent limitations that may also contribute to these outcomes.

5.1. Potential Applications and Future Directions

The Electrochemistry Assistant holds substantial potential to transform how information is accessed and utilized within the field of electrochemistry. In education, it can be integrated into learning platforms to provide students with interactive experiences, addressing their inquiries, elucidating electrochemical concepts, and offering guidance on experimental procedures. For research, the assistant can expedite access to pertinent information from scholarly literature, streamline literature reviews, and support data analysis, thereby potentially accelerating the research process and fostering discoveries. Furthermore, organizations offering electrochemical products or services can leverage the system to furnish immediate responses to customer inquiries, thus enhancing customer satisfaction and potentially reducing support costs.

Future development efforts for the Electrochemistry Assistant will focus on several key areas. Primarily, the expansion of the knowledge repository through the incorporation of additional relevant documentation, such as textbooks, research articles, and industry reports, is planned to augment the system's coverage of electrochemical concepts and methodologies. Secondly, future endeavors will explore techniques to refine both the retrieval and response generation processes, including the integration of user feedback and advanced query expansion methodologies, alongside the investigation of alternative Large Language Models and embedding techniques to further enhance the accuracy and comprehensiveness of the responses. Thirdly, to augment the system's accessibility and user-friendliness, the development of an interactive user interface is planned, enabling users to seamlessly input their queries and visualize the retrieved information.

By addressing these limitations and pursuing these future directions, the aim is to further enhance the Electrochemistry Assistant's capabilities and establish it as a more valuable resource for

the electrochemistry community. It is posited that RAG-based systems such as the Electrochemistry Assistant possess considerable potential to transform the modalities through which information is accessed and utilized within scientific domains.

6. Conclusion

In conclusion, this paper has detailed the development and evaluation of the Electrochemistry Assistant, a Retrieval-Augmented Generation system that utilizes the Gemini model and ChromaDB for electrochemical information retrieval. The evaluation results demonstrate the system's efficacy in providing precise and contextually significant responses to a defined set of test questions. These findings underscore the potential of the RAG methodology for constructing specialized question-answering systems within scientific domains like electrochemistry, offering an effective approach for accessing and synthesizing information from technical documentation. Future work will focus on expanding the knowledge corpus, refining the system's capabilities, and developing an intuitive user interface to improve accessibility, thereby further enhancing its utility within this specialized field.

Author Contributions: A.N. conceived and designed the study, developed the software implementation, and contributed to the drafting of the manuscript. P.S. led the preparation of the manuscript. Both authors provided critical feedback on the design, data analysis, interpretation of results, and the content of the manuscript.

Conflicts of Interest: The authors declare that no conflict of interest exists.

Supplementary Information: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. The complete software code, associated datasets, and further detailed information are accessible in the project repository located at: <https://github.com/anatarajank/Electrochemistry-Assistant>. This repository is made available under the terms of the Creative Commons Attribution 4.0 International License (CC-BY-4.0 International), which permits the unrestricted use, distribution, and adaptation of the work, provided that appropriate credit is given to the original authors. Moreover, the results of the system performance evaluation have been provided in the supplementary information.

References

1. Bard, A. J.; Faulkner, L. R. *Electrochemical Methods: Fundamental and Applications*, 2nd ed.; John Wiley & Sons, Inc., 2001.
2. Kissinger, Peter.; Heineman, W. R. . *Laboratory Techniques in Electroanalytical Chemistry, Revised and Expanded*; Kissinger, P. T., Heinema, W. R., Eds.; CRC Press, 2018. <https://doi.org/10.1201/9781315274263>.
3. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Kütter, H.; Lewis, M.; Yih, W. T.; Rocktäschel, T.; Riedel, S.; Kiela, D. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Adv Neural Inf Process Syst* **2020**, 2020-December.
4. Karpukhin, V.; Oğuz, B.; Min, S.; Lewis, P.; Wu, L.; Edunov, S.; Chen, D.; Yih, W. T. Dense Passage Retrieval for Open-Domain Question Answering. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* **2020**, 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>.
5. Izacard, G.; Grave, E. Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering. *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference* **2020**, 874–880. <https://doi.org/10.18653/v1/2021.eacl-main.74>.
6. Agarwal, S.; Laradji, I. H.; Charlin, L.; Pal, C. LitLLM: A Toolkit for Scientific Literature Review. **2024**.
7. Han, B.; Susnjak, T.; Mathrani, A. Automating Systematic Literature Reviews with Retrieval-Augmented Generation: A Comprehensive Overview. *Applied Sciences* **2024**, Vol. 14, Page 9103 **2024**, 14 (19), 9103. <https://doi.org/10.3390/APP14199103>.
8. Aytar, A. Y.; Kilic, K.; Kaya, K. A Retrieval-Augmented Generation Framework for Academic Literature Navigation in Data Science. **2024**.
9. Xu, Z.; Cruz, M. J.; Guevara, M.; Wang, T.; Deshpande, M.; Wang, X.; Li, Z. Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering. *SIGIR 2024 - Proceedings*

of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval **2024**, 2905–2909. <https://doi.org/10.1145/3626772.3661370>.

- 10. Zhou, X.; Zhang, S.; Agarwal, M.; Akroyd, J.; Mosbach, S.; Kraft, M. Marie and BERT—A Knowledge Graph Embedding Based Question Answering System for Chemistry. *ACS Omega* **2023**, *8* (36), 33039–33057. https://doi.org/10.1021/ACSOMEGA.3C05114/ASSET/IMAGES/MEDIUM/AO3C05114_0019.GIF.
- 11. Wellawatte, G. P.; Guo, H.; Lederbauer, M.; Borisova, A.; Hart, M.; Brucka, M.; Schwaller, P. ChemLit-QA: A Human Evaluated Dataset for Chemistry RAG Tasks. *Mach Learn Sci Technol* **2025**, *6* (2), 020601. <https://doi.org/10.1088/2632-2153/ADC2D6>.
- 12. Tran, D.; Pascazio, L.; Akroyd, J.; Mosbach, S.; Kraft, M. Leveraging Text-to-Text Pretrained Language Models for Question Answering in Chemistry. *ACS Omega* **2024**, *9* (12), 13883–13896. https://doi.org/10.1021/ACSOMEGA.3C08842/ASSET/IMAGES/LARGE/AO3C08842_0006.JPG.
- 13. Wang, C.; Yang, Y.; Song, J.; Nan, X. Research Progresses and Applications of Knowledge Graph Embedding Technique in Chemistry. *J Chem Inf Model* **2024**. https://doi.org/10.1021/ACS.JCIM.4C00791/SUPPL_FILE/CI4C00791_SI_001.PDF.
- 14. Zhou, X.; Nurkowski, D.; Menon, A.; Akroyd, J.; Mosbach, S.; Kraft, M. Question Answering System for Chemistry—A Semantic Agent Extension. *Digital Chemical Engineering* **2022**, *3*, 100032. <https://doi.org/10.1016/J.DCHE.2022.100032>.
- 15. Zhou, X.; Nurkowski, D.; Mosbach, S.; Akroyd, J.; Kraft, M. Question Answering System for Chemistry. *J Chem Inf Model* **2021**, *61* (8), 3868–3880. https://doi.org/10.1021/ACS.JCIM.1C00275/ASSET/IMAGES/MEDIUM/CI1C00275_0009.GIF.
- 16. Zhou, X.; Nurkowski, D.; Menon, A.; Akroyd, J.; Mosbach, S.; Kraft, M. Question Answering System for Chemistry—A Semantic Agent Extension. *Digital Chemical Engineering* **2022**, *3*, 100032. <https://doi.org/10.1016/J.DCHE.2022.100032>.
- 17. Chen, K.; Lu, J.; Li, J.; Yang, X.; Du, Y.; Wang, K.; Shi, Q.; Yu, J.; Li, L.; Qiu, J.; Pan, J.; Huang, Y.; Fang, Q.; Heng, P. A.; Chen, G. Chemist-X: Large Language Model-Empowered Agent for Reaction Condition Recommendation in Chemical Synthesis. **2023**.
- 18. Chukwu, R. *Home - EISight*. <https://chat.fitmyeis.com/> (accessed 2025-05-16).
- 19. Natarajan, A. *Electrochemistry Assistant*. <https://www.kaggle.com/code/aravindannatarajan/electrochemistry-assistant> (accessed 2025-05-16).
- 20. McD, M. *Day 1 - Prompting*. <https://www.kaggle.com/code/markishere/day-1-prompting#Day-1---Prompting> (accessed 2025-05-16).
- 21. Natarajan, A. *anatarajank/Electrochemistry-Assistant*. <https://github.com/anatarajank/Electrochemistry-Assistant> (accessed 2025-05-19).
- 22. Kelly, R. S. *Analytical electrochemistry, basic concepts*. https://www.asdl.org/onlineArticles/ecourseware/Kelly_Potentiometry/EC_CONCEPTS1.HTM (accessed 2025-05-16).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.