

Automated qEEG Case Study Generation with Retrieval-Augmented AI and Clinical Data Integration

Netanel Stern*

Independent Research Author, Israel

*Corresponding Author

Netanel Stern, Independent Research Author, Israel.

Submitted: 2025, Oct 03; Accepted: 2025, Oct 28; Published: 2025, Nov 06

Citation: Stern, N. (2025). Automated qEEG Case Study Generation with Retrieval-Augmented AI and Clinical Data Integration. *World J Radiolo Img*, 4(1), 01-04.

Abstract

Quantitative electroencephalography (qEEG) offers objective biomarkers of brain function across neuropsychiatric conditions, but clinical EEG case reports are traditionally labor-intensive to produce. We describe a reproducible Python-based pipeline that automatically processes raw BrainVision EEG data, extracts spectral qEEG features, integrates patient clinical scores (e.g. Brief Psychiatric Rating Scale, BPRS), retrieves relevant literature via Europe PMC, and uses a retrieval-augmented large language model (RAG-LLM) to generate structured narrative case reports. EEG preprocessing (filtering, artifact removal, referencing) and feature computation (power in delta, theta, alpha, beta bands, etc.) are implemented using open-source MNE-Python tools in a BIDS-compliant framework [1,2]. Patient metadata such as age, diagnosis, and BPRS severity provide clinical context alongside EEG features (e.g. the known increase in theta power and theta/beta ratio in schizophrenia [3]). Key EEG findings are combined with dynamically retrieved evidence from Europe PMC – an openaccess repository of ~36 million biomedical abstracts and 5 million full-text articles– to ground the report in up-to-date knowledge. Using a RAG-LLM approach, the system formulates context-aware prompts that guide the model to cite recent studies and summarize findings [4]. For example, prior work has shown retrieval-augmented LLMs significantly improve accuracy in clinical question answering compared to base models, and in dedicated frameworks (e.g. EEG-MedRAG) unify EEG domain knowledge and patient data for diagnostic guidance [5,6]. Our pipeline yields a draft case report that mimics the structure of a clinician's report: background, methods, results (EEG summary and clinical scores), and an evidence-supported discussion.

1. Introduction

EEG remains an indispensable tool in neuroscience and psychiatry, providing noninvasive recordings of brain activity. Quantitative EEG (qEEG) – the analysis of EEG power spectrum bands – has been studied as a putative biomarker in disorders such as schizophrenia, ADHD, depression and bipolar disorder [3,7]. For example, schizophrenia is often associated with increased delta/theta and reduced alpha power [3]. Clinical context is captured by symptom scales like the BPRS, which quantify severity of psychiatric symptoms and are routinely collected in research studies [3]. Integrating EEG features with clinical metrics can enhance interpretation (e.g. correlating theta increase with BPRS depression subscore). However, manually generating a cohesive case report that synthesizes EEG analyses with relevant literature is time-consuming and subjective.

Recent advances in large language models (LLMs) and retrieval-augmented generation (RAG) offer a new paradigm: an AI-driven pipeline can automatically assemble multimodal data and external knowledge into an explanatory narrative. LLMs have demonstrated capability in medical domains, but they are prone to hallucination unless grounded by real data [5,8]. RAG addresses this by coupling an LLM with a knowledge base: the model retrieves pertinent documents (here from Europe PMC) and conditions its output on this evidence [9]. Studies in healthcare have shown that RAG-augmented systems yield more accurate, up-to-date answers than base LLMs alone [5,8]. For instance, Masannek *et al.*

Tested multiple LLMs on neurology guidelines and found that a fixed-document RAG setup markedly improved accuracy over

unfettered models, though caution remains for hallucinations and case-based scenarios [5]. Similarly, Kuo *et al.* describe a hierarchical RAG pipeline that retrieves heterogeneous clinical trial data and generates reports with higher factual consistency and greatly reduced authoring time compared to manual methods [8,10]. Inspired by such successes, we propose applying RAG to qEEG.

Figure 1 illustrates the overall workflow. Raw EEG (BrainVision format) and clinical data enter the pipeline, undergo automated preprocessing and feature extraction, then trigger targeted literature search. Retrieved documents inform a final LLM prompt that produces the narrative case report.

2. EEG Data Preprocessing and Spectral Feature Extraction

Raw EEG data (e.g. BrainVision *.eeg/.vhdr* files) are ingested and organized according to the Brain Imaging Data Structure (BIDS) for neurophysiology [1]. We use MNE-Python to apply a standard preprocessing chain: band-pass filtering (e.g. 1–50 Hz), removal of line noise, and artifact correction (automatic identification of bad channels, Independent Component Analysis for ocular/muscle artifacts) [2]. It is critical to document each step for reproducibility: we leverage the MNE-BIDS-Pipeline framework, which provides scripted execution of preprocessing steps with caching and provenance tracking [1,11].

This ensures that the exact filtering, referencing (common average or linkedmastoids), and artifact-rejection parameters are recorded for audit and reuse. Recent work emphasizes that even such preprocessing choices can dramatically affect downstream analyses so consistency is vital in a clinical research context [12]. From the cleaned continuous EEG, we compute spectral power in canonical bands (delta 1–4 Hz, theta 4–8 Hz, alpha 8–13 Hz, beta 13–30 Hz, etc.) using Welch’s method or multitapering. Relative band power and ratios (e.g. theta/beta) are computed per channel and averaged over regions of interest. These qEEG features are saved in a structured format (CSV or JSON) along with metadata such as channel montages and patient demographics.

This numeric summary forms the quantitative core of the report (for example, “Global theta power was elevated to 150% of the normative mean, consistent with prior findings in schizophrenia 3”). We also compute EEG complexity or connectivity metrics (entropy, coherence) as advanced optional features. Importantly, the feature extraction is coded in Python using open-source libraries (MNE, SciPy), and the entire pipeline from raw data to feature table can be re-run end-to-end, fulfilling reproducibility standards in bioengineering [1,2].

3. Clinical Data Integration

In addition to EEG, we integrate patient-specific clinical information to contextualize findings. For example, we include diagnosis, medication status, age, and standardized scores such as the BPRS (for psychiatric symptoms), HAM-D (for depression), or MoCA (for cognition). These data may come from an electronic health record or study database. In our framework, clinical scores

are merged with EEG results so that the LLM can mention them (e.g. “The patient’s BPRS score was 28, indicating moderate schizophrenia symptoms [3]).

Prior studies often correlate EEG power changes with symptom scales. As one example, Newson & Thiagarajan note that schizophrenia severity was assessed using PANSS and BPRS in most EEG studies [3]. By including such measures, the generated report can explain how EEG abnormalities align (or do not align) with clinical severity. This multimodal integration also allows RAG to retrieve literature linking EEG markers and clinical metrics. For instance, a search combining “theta power schizophrenia BPRS” may yield studies discussing EEG predictors of symptom improvement, which the narrative can cite.

4. Literature Retrieval and RAG-Based Report Generation

To produce an evidence-based narrative, the system queries Europe PMC for relevant literature. Europe PMC is an open-access life sciences repository containing ~36 million article abstracts and 5 million fulltexts [4]. We construct search queries using patient context (e.g. “schizophrenia EEG theta”), methodological terms (e.g. “qEEG spectrum analysis software”), and any novel findings (e.g. “delta power increase clinical meaning”). Using the Europe PMC RESTful API (or associated Python libraries), the pipeline retrieves top-ranking abstracts and open-access full-texts matching these queries. The selection is filtered for recency and relevance (for example, the last 10 years, human studies, English language).

The core of report generation uses a Retrieval-Augmented Generation model. Retrieved documents (titles, snippets, or passages) form an evidence bank. We then prompt a large language model (e.g. GPT-4 or a fine-tuned domain model) with both the structured patient/EEG data and key excerpts from the literature. The prompt instructs the LLM to write a structured report, ensuring each statement is grounded in the retrieved evidence. For example, in the LLM prompt we include: “*Patient is a 30-year-old with schizophrenia (BPRS 30) whose EEG shows elevated theta power (mean 8 μV^2). According to [Smith et al. 2022], increased theta power correlates with positive symptoms. Summarize these findings in a report with references.*”

The RAG approach has demonstrated improved factual accuracy in medical summaries [8,9]. Our application is analogous to systems like AlzheimerRAG, which fuse textual and imaging data for case studies [13], and the EEG-MedRAG framework which builds hypergraphs of EEG knowledge and patient data for causal diagnosis generation [6]. The output is a draft report comprising sections: **Background** (patient demographics, clinical history), **EEG Acquisition** (recording details, preprocessing), **Results** (qEEG features with normative comparisons), and **Discussion** (interpretation citing literature).

In Discussion, the LLM references specific studies (e.g. “The observed theta increase aligns with reports of frontal slowing in schizophrenia3”) and notes if findings contradict literature (e.g. no alpha slowing despite expectation). Each cited fact is traced

to a reference from Europe PMC to maintain transparency. The language model is steered to a “report-writing” style, using sentence templates extracted from sample case studies.

Figure 2 outlines the retrieval-to-generation flow: EEG features and patient data formulate a multi-part prompt; the RAG system retrieves abstracts; the LLM synthesizes a prose report with in-text citations to the sources.

5. Reproducibility and Open Pipeline Implementation

A key design goal is full reproducibility. All code is in Python and managed with version control. The data flow is modular: BIDS validation ensures input conformity, MNE/BIDS-Pipeline handles preprocessing with cacheable steps and feature computation scripts log their parameters [1,11]. We containerize the environment (e.g. with Docker or Conda) so others can recreate the exact software setup. The RAG component is also documented: the retrieval queries and LLM prompts are saved alongside the results. This allows independent verification of the report content.

To facilitate reuse by researchers, we leverage community standards. Using BIDS format means that any EEG dataset following BIDS-EEG can be plugged into the pipeline [1]. We provide example Jupyter notebooks that walk through each step on sample data. The pipeline can run in parallel for multiple subjects, supporting large-scale studies (as advertised by the MNE-BIDS-Pipeline for hundreds of datasets [11]). Summaries of processing (filter logs, artifact rejection rates, feature distributions) are automatically compiled into a report PDF, enabling quick quality checks. By being open-source, this framework advances the ethos of reproducible neuroengineering practice.

6. Clinical Research Utility, Education, and Future Directions

This automated reporting tool has several utilities. In clinical research, it accelerates the generation of case studies and cohort summaries. Investigators can use it to standardize EEG report content across studies, reducing variability. The inclusion of RAG ensures that reports cite current literature, keeping interpretations up to date, which is crucial in fields where biomarker validity is evolving. In neuroengineering education, the pipeline serves as a teaching aid: students can explore how preprocessing choices affect features, and how AI can assist in interpreting neurophysiological data [12]. The system can generate example cases for training, highlighting how spectral changes relate to diagnosis and literature.

Looking forward, the framework can be extended. Future versions might incorporate other modalities (e.g. MRI or genetics) into the RAG context, enabling truly multimodal case reports. Improving the LLM’s domain specificity (through fine-tuning on neuroengineering literature) could reduce errors. There is also potential for real-time use: integrating with EEG acquisition software to update reports as data are collected. From a bioengineering perspective, such tools illustrate how AI can bridge raw data and clinical insight, embodying the translational promise of neuroinformatics.

7. Strengths, Limitations, and Outlook

This approach offers major strengths: automation greatly reduces expert time, promotes consistency, and ties findings to evidence [5,8]. By using open pipelines and data standards, it encourages reproducible research and democratizes complex EEG analysis for non-experts. However, limitations remain. EEG preprocessing is sensitive; suboptimal filtering or artifact correction can mislead analysis [12]. The quality of the LLM report hinges on retrieval: if relevant literature is missed or irrelevant documents are retrieved, the narrative may be skewed or incomplete. As noted in prior RAG studies, LLMs can still hallucinate or oversimplify in clinical contexts so reports must be reviewed by experts [5,8]. Data privacy is also a concern: patient data used in prompts should be de-identified and handled under appropriate governance. In future work, evaluation is critical: we plan systematic testing of report accuracy by comparing AI-generated reports with those by neurophysiologists. Advances in domain-specific LLMs and larger EEGtext corpora will likely improve performance. Overall, merging automated EEG analytics with retrieval-augmented AI represents a promising direction in bioengineering — one that could transform how we synthesize physiological data, clinical scores, and biomedical knowledge into actionable insights.

Figure 1 | Schematic of the automated qEEG case report pipeline. Raw EEG (BrainVision) and clinical data (demographics, BPRS, etc.) are preprocessed (filtering, artifact removal) in a BIDS/MNE-Python workflow. Spectral features (band powers, ratios) are extracted and combined with patient context. A query generator then submits relevant keywords to the Europe PMC API to retrieve literature. Finally, a retrieval-augmented LLM (e.g. GPT-4) is prompted with the data and retrieved evidence to produce the narrative report (automatically formatted with sections and citations).

Figure 2 | Flowchart of RAG-based report generation. Patient EEG features and metadata are used to construct search queries. Retrieved abstracts/passages from Europe PMC form the knowledge base. A large language model is then conditioned on both the structured data and the retrieved text to generate a factual, citation-supported EEG case study report.

References

1. Newson, J. J., & Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: a review of resting state studies. *Frontiers in human neuroscience*, 12, 521.
2. Kessler, R., Enge, A., & Skeide, M. A. (2025). How EEG preprocessing shapes decoding performance. *Communications Biology*, 8(1), 1039.
3. Wang, Y., Luo, H., & Meng, L. (2025). EEG-MedRAG: Enhancing EEG-based Clinical Decision-Making via Hierarchical Hypergraph Retrieval-Augmented Generation. *arXiv preprint arXiv:2508.13735*.
4. Masannek, L., Meuth, S. G., & Pawlitzki, M. (2025). Evaluating base and retrieval augmented LLMs with document or online support for evidence based neurology. *npi Digital Medicine*, 8(1), 137.

-
5. Kuo, S. M., Tai, S. K., Lin, H. Y., & Chen, R. C. (2025). Automated Clinical Trial Data Analysis and Report Generation by Integrating Retrieval-Augmented Generation (RAG) and Large Language Model (LLM) Technologies. *AI*, 6(8), 188.
 6. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroinformatics*, 7, 267.
 7. EMBL-EBI Literature Services. Europe PMC database and text mining infrastructure (2025).

Copyright: ©2025 Netanel Stern. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.