Submission code: 7W9BKMP7YG

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Seismic Data Analysis with AI Agents and Large Language Models

José Augusto Agripino de Oliveira (Universidade Federal do Rio Grande do Norte), Victor do Nascimento Gomes (Universidade Federal do Rio Grande do Norte), Ernane Ferreira Rocha Junior (Universidade Federal do Rio Grande do Norte), Ignacio Sánchez-Gendriz (Universidade Federal do Rio Grande do Norte), Luiz Affonso Henderson Guedes De Oliveira (Universidade Federal do Rio Grande do Norte), Ivanovitch Medeiros Dantas Da Silva (Universidade Federal do Rio Grande do Norte), Cristhian Alberto Celestino Cortez (Cenpes-Petrobras), Pablo Machado Barros (Cenpes-Petrobras), Tiago Ilipronti Girardi (Cenpes-Petrobras), Luis Fernando Mendes Cury (Cenpes-Petrobras), Leandro da Silva Sadala Valente (Cenpes-Petrobras), Lucas Gondim Miranda (Cenpes-Petrobras), Tiago Barros (Universidade Federal do Rio Grande do Norte)

SBGf Conference
18-20 NOV | Rio'25

Seismic Data Analysis with AI Agents and Large Language Models

Abstract Summary

Seismic data exploration and analysis pose significant challenges due to the data's inherent complexity, volume, and heterogeneity. Identifying relevant seismic datasets for processing remains time-consuming due to the vast number of available files. Traditionally, preliminary analysis requires specialized software or programming expertise to assess whether the data merits further processing, resulting in increased costs and delays. Recent advances in Artificial Intelligence (AI), Machine Learning (ML), Large Language Models (LLMs), and agent-based systems have opened new possibilities for addressing these challenges. In this context, we present the Brazilian Intelligence for Seismic Analysis (BRISA), a system designed to facilitate the initial inspection of SEG-Y files by enabling interaction without requiring programming skills or dedicated software. BRISA allows users to interact with seismic data through natural language, offering a user-friendly interface that supports the visualization of header information and acquired data. The system generates interactive plots that enhance exploratory analysis and allow dynamic selection among various LLMs, including GPT, LLaMA, and Gemini models. By simplifying complex operations through intuitive commands, BRISA improves the accessibility of seismic data analysis and reduces the technical barriers typically associated with the initial data examination phase.

Introduction

Seismic datasets play a key role in scientific research and industrial geophysics by providing valuable insights into the Earth's subsurface and related processes. However, their sheer volume, complex structure, and heterogeneity of seismic datasets make analysis and interpretation challenging (Khosro Anjom et al., 2024). In addition to the signal traces, these datasets include necessary metadata (Davis and Hunt, 2024). For example, SEG-Y (Norris and Faichney, 2002) file headers store essential contextual information, like source and receiver coordinates, trace numbers, offsets, and acquisition parameters, that are crucial for accurate spatial analysis and data processing.

Conventional approaches to seismic analysis often rely on proprietary software suites that require advanced programming skills and significant computational resources, thereby limiting accessibility for many researchers and professionals. Despite growing interest in AI-assisted geophysical analysis, few tools enable natural language interaction with seismic data or simplify early inspection tasks for non-specialists. To address this issue, BRISA was developed using AI agents, LLMs, and tools built with modern libraries. This work presents a solution that enables users to interact with seismic data for exploration and visualization through natural language. This significantly streamlines the initial analysis phase, saving time and resources so that users can identify the most relevant data before proceeding with complete seismic processing.

The agent-based system reduces reliance on proprietary software and programming expertise for preliminary seismic data analysis, facilitating this crucial step and helping to mitigate resource waste. The solution supports integration with LLMs and tools developed using open-source libraries,

operating through agents. Key features include choosing the LLM to be used, conversation logging, visualization of the agent's execution flow, and interaction with seismic data and SEG-Y header information. Dynamic graphs are generated from the data, allowing users to zoom in, zoom out, and export the graph as a PNG file.

Theory and Method

Large Language Models (LLMs) represent a category of deep learning architectures trained on extensive text datasets to learn statistical and semantic patterns in natural language (Xavier et al., 2024; Zhang et al., 2025). LLMs are built on transformer architectures (Vaswani et al., 2017) and enable context-sensitive text generation and interpretation, exhibiting generalization capabilities comparable to human performance in tasks such as summarization, code generation, and specialized reasoning (Ilić and Gignac, 2024).

An agent is a system that leverages LLMs to interact with its environment through user instructions. It combines reasoning, planning, and executing actions, often using external tools, to complete tasks (Burtenshaw et al., 2025). Tools are extra functions that extend the LLM capabilities, such as data manipulation, parsing trace headers, plotting acquisition geometry, or generating visualizations of selected seismic attributes. Orchestration frameworks such as LangChain (Chase, 2022) and LangGraph (Inc, 2024) manage the integration between agents and tools. These frameworks define and manage the interaction logic among LLMs, agents, and tools, supporting flexible and scalable workflows for seismic data analysis. By combining the reasoning capacity of LLMs with the operational precision of tool execution, intelligent agents enable complex and customizable geophysical analyses to be conducted through natural language interaction.

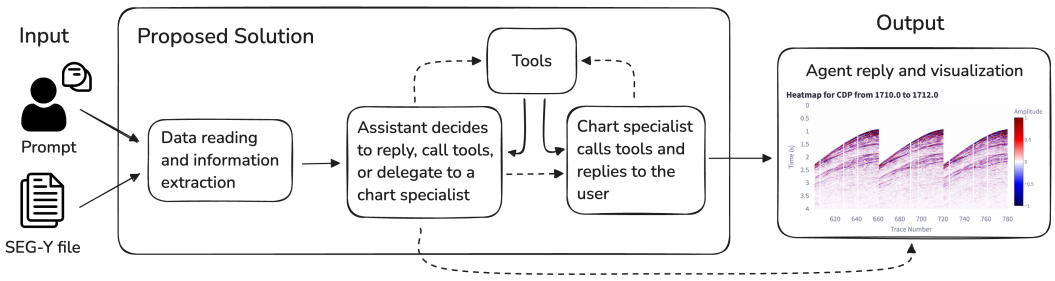

Figure 1: Flowchart of the BRISA system. The pipeline processes SEG-Y files and delegates tasks to specialized agents for data queries, visualizations, and natural language interaction.

Figure 1 presents a flow diagram of the LangGraph workflow, highlighting its modular architecture for user interaction, data exploration, and visualization. The proposed solution takes the user command and the SEG-Y file as input. It starts by reading and extracting information from the data using the Segyio (Equinor, 2025) and Segy-SAK (Hallam, 2025) libraries. Then, an agent decides whether to respond directly to the user's prompt, call external tools, or delegate the task to a specialized chat agent. If the user requests a graphic, the assistant delegates the task to a chart specialist agent, which uses specific tools to generate the chart and compose an appropriate response. Each agent has its own set of tools, which are not shared. The first agent handles general responses and data manipulation, while the second focuses on chart generation. The final output is a response based on the context of the conversation and tool interactions. Any generated chart is displayed using the Streamlit (Inc., 2019) interface and is interactive, thanks to the Plotly (Inc., 2015) library.

Results

The developed system presents an intuitive interface that allows users to choose which LLM model to use, update SEG-Y files, save chats, and visualize the LangGraph workflow (Figure 2). This functionality was designed to facilitate advanced seismic data analysis and interactive exploration of SEG-Y files using natural language commands.

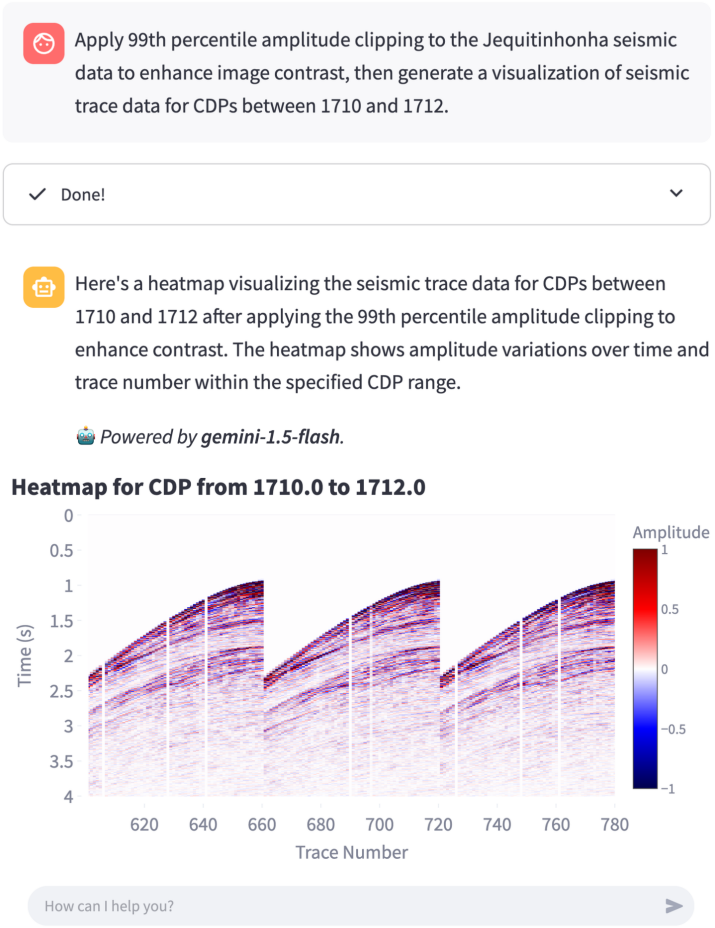

Figure 2: Visualization of seismic trace data with 99th percentile amplitude clipping for CDPs from 1710 to 1712. The X-axis is 'Trace Number', the Y-axis (increasing downward) is 'Time (s)', and the color represents the amplitude of the captured signal.

Using the Segyio library, it is possible to represent a SEG-Y file as a DataFrame-like structure and perform preprocessing on both the SEG-Y header metadata and the seismic trace data, according to the user's prompt and the available tools. Users can sort seismic trace data based on primary and secondary keys present in the file header, apply amplitude clipping using a percentile threshold, and scale the data by normalizing the values of source and receiver coordinates.

Integrating interactive visualization libraries like Streamlit and Plotly significantly enhances the analytical workflow by enabling real-time data interaction and visualization. Through the system's natural language interface, users can generate wiggle plots, histograms, scatter plots, line charts,

heatmaps, and other visualization types. Figure 2 illustrates a user command that first applies 99th percentile amplitude clipping to the Jequitinhonha seismic dataset to enhance image contrast—a common practice for improving data visualization and interpretability. The command also requests a plot of the seismic data for CDPs ranging from 1710 to 1712.

Conclusions

This study demonstrates the feasibility and practicality of using an LLM-powered agent system with specialized tools to interact with seismic data through natural language. Leveraging recent advances in large language models, orchestration frameworks like LangChain and LangGraph, and libraries like Streamlit and Plotly, BRISA enables users to perform complex tasks without writing code or relying on proprietary software. Users can sort seismic traces by header keys, clip amplitudes using percentile thresholds, normalize coordinates, and generate various visualizations—such as wiggle plots, histograms, scatter plots, and heatmaps—using plain language instructions. This lowers the entry barrier for non-specialists and accelerates workflows for experienced users. By transforming seismic data handling into an intuitive, language-driven process, BRISA streamlines early-stage analysis and helps users identify which SEG-Y files are worth further processing. Future enhancements include support for additional data formats and deployment in real-world operational environments.

References

Burtenshaw, B., J. Thomas, and T. Simonini, 2025, The hugging face agents course: `https://github.com/huggingface/agents-course`. (GitHub repository).

Chase, H., 2022, Langchain.

Davis, W., and C. R. Hunt, 2024, Knowledge graphs for seismic data and metadata: Applied Computing and Geosciences, **21**, 100151.

Equinor, 2025, Segyio: Fast python library for seg-y files. (Accessed: April 2025).

Hallam, A., 2025, Segy-sak. (Accessed: April 2025).

Ilić, D., and G. E. Gignac, 2024, Evidence of interrelated cognitive-like capabilities in large language models: Indications of artificial general intelligence or achievement?: Intelligence, **106**, 101858.

Inc, L., 2024, Langgraph.

Inc., P. T., 2015, Collaborative data science.

Inc., S., 2019, Streamlit: Turn data scripts into shareable web apps. (Open source framework for creating interactive data applications.).

Khosro Anjom, F., F. Vaccarino, and L. V. Socco, 2024, Machine learning for seismic exploration: Where are we and how far are we from the holy grail?: GEOPHYSICS, **89**, WA157–WA178.

Norris, M. W., and A. K. Faichney, 2002, Seg y rev 1 data exchange format. (Accessed: April 2025).

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, 2017, Attention is all you need: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., 6000–6010.

Xavier, M., T. Laikh, S. Patil, and V. Vyatkin, 2024, Llm-powered multi-actor system for intelligent analysis and visualization of iec 61499 control systems: IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 1–8.

Zhang, J., C. Clairmont, X. Que, W. Li, W. Chen, C. Li, and X. Ma, 2025, Streamlining geoscience data analysis with an llm-driven workflow: Applied Computing and Geosciences, **25**, 100218.