https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

GeoSIRR 1.0: Conversational Geological Cross-Section Modeling Using Large Language Models

Denis Anikiev[1], Juan Esteban Mosquera[2], Korhan Ayranci[3], Judith Bott[4], and Umair Bin Waheed[3]

[1]Center for Integrative Petroleum Research, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
[2]Geophysics Group, Universidad Nacional de Colombia, Bogotá, Colombia
[3]Department of Geosciences, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
[4]Section 4.5 – Subsurface Process Modelling, GFZ Helmholtz Centre for Geosciences, Potsdam, Germany

Correspondence: Denis Anikiev (denis.anikiev@kfupm.edu.sa)

Abstract. Geological cross-sections are a fundamental tool for subsurface interpretation, yet their construction remains a labor-intensive and largely manual process that relies on expert judgment and structured numerical inputs. While recent advances in artificial intelligence have enhanced specific geoscientific workflows, no existing method enables the direct generation and iterative refinement of geological cross-sections from unstructured natural language descriptions. In this paper, we present GeoSIRR 1.0 (Geological Section Interpretation, Reconstruction & Refinement), a novel modeling framework that leverages large language models (LLMs) to translate free-form geological narratives into structured, coordinate-based cross-section geometries. GeoSIRR introduces a domain-specific language (DSL) for representing geological bodies as topologically consistent polygons and integrates automated geometric and geological validation to ensure continuity, stratigraphic consistency, and structural plausibility. The framework supports both initial model generation and conversational refinement, allowing users to iteratively modify cross-sections using natural language commands while preserving existing geometry. We demonstrate the capabilities of GeoSIRR through multiple geological scenarios, including faulted sedimentary systems, intrusive bodies, and progradational deltaic sequences, and assess repeatability across multiple generation runs. Results show that GeoSIRR consistently produces geologically plausible cross-sections and effectively incorporates conceptual refinements with reduced generation time compared to initial model construction. By directly linking qualitative geological reasoning with quantitative geometric modeling, GeoSIRR provides a self-contained, dialogue-driven approach to cross-section construction that complements existing modeling tools and offers new opportunities for education, exploratory analysis, and rapid scenario development in subsurface geoscience.

1 Introduction

Geological modeling is a fundamental and essential process in subsurface characterization, critical for applications spanning petroleum exploration and production, mineral resource assessment, groundwater management, geotechnical engineering, and carbon storage. The ability to accurately represent subsurface stratigraphic architecture and structural geology components directly impacts the quality of exploration decisions, risk assessment, and economic viability of development projects (Zakrevsky, 2011). Although there are several ways to represent subsurface architectures (e.g., wireline log correlation, and seismic data

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

interpretation), modeling is the most commonly used, probably because it allows users to test various scenarios relatively fast (e.g., Cao et al., 2024).

Typical geological modeling workflows often include manual two-dimensional cross-section interpretation, which serves as the foundation for many subsurface studies and can be rapidly prototyped and iteratively refined based on available data (Caumon et al., 2009). Three-dimensional volumetric models represent a natural extension, providing comprehensive spatial representation of subsurface heterogeneity across basins. Block-model representations, commonly used in mining and resource assessment, discretize the subsurface into regular volumes with associated properties (Zakrevsky, 2011; Bonham-Carter, 2016). More advanced approaches include implicit surface modeling, which uses mathematical interpolation functions to define geological boundaries continuously in space, and process-based forward models, which simulate sedimentary and structural processes to predict realistic geological architectures (Pyrcz and Deutsch, 2014; Zhong et al., 2019).

The choice of modeling approach depends on project objectives, available data quality and density, computational resources, time constraints, and the geoscientific team's expertise. Cross-section modeling, despite its apparent simplicity, is a foundational tool in geoscience for visualizing and interpreting the Earth's subsurface architecture. A geological cross-section provides a vertical slice through the subsurface, allowing geoscientists to clarify relationships between rock layers, faults, and other features which are the information for petroleum reservoir characterization, mineral exploration, and geotechnical site assessment (Boggs, 2006; Fossen, 2016).

Conventional approaches to geological modeling, however, face significant limitations. In practice, cross-sections are often constructed manually by expert geologists who correlate limited field observations (e.g., wireline logs and outcrop data) along a profile and infer the geology between known data points (Caumon et al., 2009; Bond, 2015). This process is labor-intensive and inherently subjective, relying on the geologist's experience to draw unit boundaries and satisfy stratigraphic principles. On the other hand, a variety of computer-aided tools implement rule-based or algorithmic methods to generate models (de la Varga et al., 2024; Grose et al., 2021). These include explicit "connect-the-dots" techniques and implicit modeling algorithms that interpolate surfaces through data points based on mathematical constraints (Hassanzadeh et al., 2022).

While such tools can accelerate model building and allow testing different scenarios, they typically require structured numerical inputs, significant uncertainties and predefined rules, making it difficult to incorporate nuanced expert insights. In particular, translating qualitative geological knowledge (e.g., a geologist's description of how a formation pinches out or a conceptual depositional history) into the rigid parameters of a model remains challenging (Chan and Elsheikh, 2017). Traditional modeling software also tends to lack intuitive, interactive interfaces: adjusting a model often involves manual parameter tuning or scripting, which may not be a geologist-friendly or iterative process (Chen et al., 2022). These shortcomings hinder the seamless integration of high-level geological understanding into models and limit the ability to refine models interactively.

Recent advances in artificial intelligence (AI), especially large language models (LLMs), offer a promising avenue to overcome these limitations. LLMs such as GPT-series models OpenAI (2023, 2025) have demonstrated an ability to comprehend and generate human-like text, including domain-specific technical descriptions. In the geological context, LLM's can act as intelligent interpreters between geological natural language description and a formal geological model. With appropriate prompt engineering and domain context, these models are capable of parsing abstract geological information and generating meaning-

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

ful representations (Fu et al., 2025; Lin et al., 2024). For instance, **GeoMinLM** (Fu et al., 2025) demonstrates the utility of LLMs in mineral exploration surveys by facilitating intelligent question-answering and knowledge dissemination. Similarly, **GeoGPT** (Zhang et al., 2024c) has been applied to automate complex workflows, such as constructing geological knowledge graphs and streamlining geoscience data analysis.

Leveraging an LLM's vast learned knowledge and reasoning ability, a modeling system could directly translate free-form geological narratives into meaningful subsurface cross-sections, effectively bridging the gap between qualitative expert insight and quantitative modeling results. Importantly, an LLM-driven approach would enable a more intuitive and interactive workflow: a geoscientist could describe subsurface features in plain language and iteratively refine the generated cross-section through conversation, rather than through tedious manual edits or coding. To our knowledge, no existing cross-section modeling method provides this level of direct natural language integration, which represents a novel solution to longstanding challenges in the field.

In this work, we propose **GeoSIRR** (Geological Section Interpretation, Reconstruction & Refinement), a new framework that leverages LLMs to directly convert natural language geological descriptions into structured, coordinate-based cross-section geometries. The system is able to interpet textual input to generate (reconstruct) initial cross-sections, and supports refinement through conversational commands and integrates validation mechanisms to ensure geological plausibility. The name "GeoSIRR" also evokes the Arabic word *sirr* (سِرّ), meaning "mystery" or "secret", reflecting how the system unveils the hidden complexities of subsurface geology through intuitive natural language interaction.

2 Related work

Geological modeling can be broadly categorized into conventional methods and AI-based approaches. Conventional methods primarily rely on manual interpretation, rule-based software, and numerical modeling techniques, while AI-based methods leverage advanced algorithms and machine learning techniques to enhance modeling capabilities (Karpatne et al., 2019; Dramsch, 2020).

2.1 Conventional Geological Modeling

The vast majority of the conventional geological modeling methodologies encompass both manual interpretation techniques and computer-assisted, rule-based tools. Traditionally, geologists construct cross-sections by compiling available subsurface data, such as wireline logs, field observations, and mapped outcrops, and they infer the geometry subsurface and structures between those control points. Following established stratigraphic correlation principles, they connect equivalent lithologies or marker horizons across boreholes to define unit boundaries and structural relationships (Caumon et al., 2009; Bond, 2015; Woodward et al., 1989), often employing advanced techniques like sequence stratigraphy, biostratigraphy, and chemostratigraphy (Catuneanu, 2006; Jones and Simmons, 1999; Craigie, 2018) to resolve complex depositional architectures.

This conventional method allows experts to apply their conceptual understanding (e.g., recognizing depositional trends and structural continuity) on a case-by-case basis. However, it is highly dependent on individual expertise and can be time-

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

consuming. Given the inherent sparsity of subsurface data, multiple valid geological models can typically be derived from the same dataset. While this non-uniqueness is a fundamental challenge, systematically generating and evaluating multiple alternative interpretations to bound this uncertainty is often too time-consuming when done manually. Moreover, because it is rare to directly observe all portions of a cross-section in the field, the manual approach inevitably involves extrapolation and uncertainty (Bond, 2015; Woodward et al., 1989; Wellmann and Caumon, 2018). While skilled practitioners use regional knowledge and geological reasoning to make the best possible interpretations, the manual process does not easily scale due to its time-intensive nature and reliance on limited expert availability. Furthermore, practical time constraints often force geologists to commit to a single deterministic model, preventing the systematic exploration of the full range of plausible structural and stratigraphic scenarios.

To improve efficiency and consistency, rule-based and algorithmic modeling tools have been developed. These range from classical deterministic methods (such as drawing straight-line interpolations between correlated stratigraphic contacts) (e.g., Glagolev and Petrishchevsky, 2010) to advanced software that employs implicit modeling and geostatistical algorithms (e.g., Hassanzadeh et al., 2022). For example, implicit modeling frameworks can fit smooth surfaces or volumes through sparse data points, generating continuous geological horizons automatically (Hillier et al., 2023). Other systems use knowledge-based rules (for instance, honoring stratigraphic superposition or fault displacements) to guide the construction of 3D models (Woodward et al., 1989; Jessell and Valenta, 1996). These automated techniques can rapidly produce geological cross-sections or volumes that satisfy basic geometric and stratigraphic constraints. Nevertheless, they expect structured input data and well-defined parameters which typically, users must provide cleaned datasets (borehole coordinates, surfaces, fault polygons) and sometimes tweak algorithm settings to get geologically reasonable results (Wellmann and Caumon, 2018). Such tools struggle to incorporate information that is not easily quantified. Qualitative expert knowledge, like a geologist's insight that "Formation A likely pinches out west of the fault B" or an uncertain inference drawn from regional analogs, cannot be directly fed into most modeling software. Indeed, converting qualitative geological knowledge into quantitative model inputs is noted to be a persistent challenge (Wellmann and Caumon, 2018; Jessell and Valenta, 1996; Caumon, 2010). As a result, important contextual information often gets left out of the modeling process, or must be manually translated into proxies (e.g., additional pseudo-data points or custom adjustments) by the user.

Another shortcoming of conventional modeling workflow is the lack of intuitive, interactive model refinement. With traditional software, building or editing a geological model usually involves a sequence of manual steps in a graphical user interface or editing text-based model scripts. For instance, adjusting the position of a stratigraphic boundary might require re-entering data points or altering control parameters and then re-running the interpolation. This process is not only cumbersome but also non-intuitive for conveying high-level changes. There is typically no facility for a geologist to "ask" the system to make an interpretive change (e.g., "*extend bed X further in the north-east direction based on new field evidence*" or "*adjust the dip of fault Y to be steeper in accordance with recent seismic data*") in natural language. Some modern tools provide limited interactivity, yet altering model parameters via traditional interfaces or scripting often requires translating geological intent into technical specifications (Chen et al., 2022). Streamlining this interaction to be more intuitive allows geologists to explore alternative interpretations or update the model on the fly using the same narrative logic they would use in reasoning with colleagues.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

The rigidity of these workflows can discourage iterative experimentation and makes it challenging to integrate newly acquired qualitative insights once an initial model is in place.

Recent advances in computer graphics have also improved the efficiency of conventional geological modeling. Li et al. (2025a) introduced a GPU-accelerated method for real-time generation of geological sections using per-pixel linked lists to optimize fragment storage and rendering. Their approach renders the geological model only once and reuses stored pixel-level depth and color information for rapid, multi-section visualization. Comparative experiments demonstrated substantial gains in speed and scalability compared to conventional stencil-buffer methods, enabling real-time generation of planar and curved sections even for complex models. This development illustrates how modern visualization pipelines can expand the interactivity and computational efficiency of traditional geological modeling workflows without relying on machine-learning components. While GPU-accelerated rendering techniques such as that of Li et al. (2025a) significantly improve the performance and interactivity of visualization, they still operate within the deterministic, geometry-based framework of traditional modeling. Consequently, these methods share many of the same conceptual limitations as other conventional approaches – most notably, their reliance on predefined inputs, difficulty incorporating qualitative geological knowledge, and lack of true semantic understanding of geological context.

In summary, both manual and existing computer-based geological modeling methods have notable limitations. Manual techniques, while leveraging expert intuition, are slow and inherently variable. Rule-based and algorithmic tools offer speed and consistency but demand formalized inputs and often overlook the rich qualitative context of geological knowledge. Crucially, neither approach supports a truly seamless or conversational interaction with the model. These limitations justify the need for an improved approach relying on advanced capabilities of generative AI.

2.2 AI-based Geological Modeling

Geological modeling has increasingly incorporated artificial intelligence techniques to enhance model accuracy, efficiency, and interpretability. AI-based approaches can be broadly categorized into pure machine learning methods and generative AI techniques, each contributing uniquely to the field.

2.2.1 Machine Learning Approaches

The integration of machine learning (ML) into geological modeling has led to several advances. Lyu et al. (2024) developed a ML-based model for generating 3D geological models from borehole data and trusted geological data, processing these inputs to generate multiple geological cross-sections of subsurface volumes. Their approach achieved significant improvements over traditional methods by incorporating stratigraphic pattern recognition.

Deep learning approaches for subsurface modeling have shown promising results. Abubakar et al. (2022) presented an end-to-end solution powered by multiple convolutional neural networks (CNNs) for subsurface modeling and interpretation, demonstrating high accuracy and strong lateral consistency in the Groningen gas field. Their workflow integrates data preconditioning, seismic property estimation, structural feature identification, and 3D rock property modeling.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Ren et al. (2025) proposed ET4DD (Enhanced Transformer for Drilling Data), a deep learning model that accurately predicts lithology classification of 3D points using voting mechanisms and offset-attention mechanisms. This approach precisely captures geological variability while effectively simplifying model complexity.

Hillier et al. (2023) introduced GeoINR 1.0, an implicit neural representation (MLP-based) framework that models 3-D geological structures directly from scattered observations (interfaces, unit labels, orientations). By encoding stratigraphic relations, unconformities, and smoothness as loss terms and using geometric initialization, GeoINR fits noisy/sparse data and captures complex geology across basin and deformed-terrain case studies.

These ML approaches have demonstrated the potential to enhance geological modeling by leveraging data-driven insights and advanced algorithms. However, they often require structured inputs and may not fully capture the qualitative aspects of geological knowledge, highlighting the need for more intuitive and flexible modeling frameworks.

2.2.2 Generative AI Approaches

Unlike traditional ML methods that focus on pattern recognition and predictive tasks, generative AI techniques aim to create new data instances that resemble the training data. In geological modeling, generative AI has been employed to synthesize realistic geological structures and facilitate the interpretation of complex subsurface features.

Generative Adversarial Networks (GANs) employ a dual-network architecture where a generator creates synthetic data and a discriminator evaluates its authenticity. For instance, Coiffier et al. (2020) introduced DiAGAN (Dimension Augmenter GAN), a novel approach that generates 3D geological fields from 2D training images using GAN. This method allows the creation of stochastic geological structures with sufficient resemblance and variability for geostatistical applications.

Recent efforts have started leveraging GANs to automate also the creation of geological cross-sections. For example, deep generative models have been applied to construct cross-sections from geoscience data. Ran et al. (2022) employed a conditional GAN that integrates geological maps and geophysical data to produce plausible 2D cross-sections automatically. While such data-driven approaches can capture stratigraphic complexity and structures, they do not incorporate natural language descriptions; users still must provide structured inputs such as grids or images. Automatic section generation thus remains challenging due to the requirement of significant manual drafting or rule-based interpolation, lacking a way to directly use the rich textual descriptions geologists formulate.

Variational Auto-Encoders (VAEs) are another class of generative models that learn to encode input data into a latent space and then decode it back to reconstruct the original data. In geoscientific applications, VAEs have been used to represent spatial uncertainty and geological variability within a reduced latent space, enabling efficient sampling and inversion. For instance, Lopez-Alvis et al. (2022) employed a VAE for geophysical inversion, using it to model assembled spatial priors and improve the realism of subsurface reconstructions compared to conventional deterministic approaches. Similarly, Xiong et al. (2021) proposed a physically constrained VAE for geochemical prospectivity mapping, where geological constraints were incorporated directly into the loss function to enhance interpretability and adherence to known geological processes. More recently, Huang et al. (2025) introduced a spatially constrained VAE that integrates spatial regularization into the latent space to denoise and quantify uncertainty in large-scale geochemical survey data. Together, these studies demonstrate that VAEs

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

provide an effective framework for encoding complex spatial dependencies in geological datasets while maintaining flexibility for generative reconstruction and uncertainty quantification.

Diffusion Models have emerged as a powerful class of generative models, particularly for tasks involving high-dimensional data such as images and 3D structures. These models work by simulating a diffusion process that gradually transforms a simple distribution into a complex one, allowing for the generation of realistic samples. In geological modeling, diffusion-based methods have been adopted to generate subsurface facies and reservoir architectures with improved realism and controllability. Lee et al. (2025) developed a latent diffusion model (LDM) for conditional reservoir facies generation, showing that diffusion-based approaches outperform GANs in producing geologically consistent realizations under limited data conditions. Likewise, Di Federico and Durlofsky (2025) applied an LDM framework combining a VAE and U-Net denoiser for facies-based geomodel parameterization and data assimilation, achieving high fidelity in reproducing channelized structures. Extending this line of work, Ovanger et al. (2025) conducted a statistical evaluation of diffusion models against traditional truncated Gaussian random field methods, confirming their potential for capturing non-Gaussian geological variability. These developments highlight diffusion models as a promising frontier for data-driven geological simulation, capable of generating diverse and geologically plausible realizations from limited or incomplete observations.

Flow Matching is a continuous normalizing flow-based generative method that learns to transform a simple probability distribution (e.g., Gaussian noise) into the complex distribution of target data through a learned velocity field. Instead of adversarial training (as in GANs) or iterative denoising (as in diffusion models), flow matching directly learns the gradient of the probability flow that connects the source and target distributions.

In geoscience contexts like in the Synthetic Geology project (Ghyselincks et al., 2025), flow matching allows the model to generate realistic 3D geological structures by learning a smooth transformation from random fields to geologically plausible volumes – without the instability or mode collapse issues common in GANs. This method captures a richer range of potential subsurface scenarios compared to traditional deterministic methods.

Several early investigations and prototypes in geoscience and allied disciplines have looked at using Large Language Models (LLMs) to support technical processes (Hadid et al., 2024). For instance, some research tools leverage LLMs to automate geospatial data analysis (Kuckreja et al., 2023; Zhang et al., 2024c, b; Holm, 2024; Mansourian and Oucheikh, 2024), while others provide a layer between human users and the open data API, e.g., in the field of mineralogy Zhang et al. (2024a); Fu et al. (2025), or integrate LLM into the coding and data processing of seismic full-waveform inversion (FWI) (Weijermars et al., 2023). Such approaches improve efficiency in exploratory data analysis without involving significant domain knowledge ingrained in the LLM itself.

These developments focus on particular, limited tasks instead of providing complete, flexible help across several dimensions of a challenging geoscience project. The community has recognized that general-purpose LLMs often lack specialized geoscience knowledge, prompting the creation of domain-specific models and frameworks. Several foundation models (FMs) for geosciences have emerged in the past two years (Zhang et al., 2023; Hadid et al., 2024; Liu and Ma, 2024). For instance, K2 fine-tuned a LLaMA-7B model on billions of tokens of geoscience literature to imbue it with discipline-specific terminology (Deng et al., 2024). The larger GeoGalactica (30B) model was trained on a massive corpus of 65 billion geoscience

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

tokens and further instruction-tuned on one million Q&A pairs, achieving impressive performance on knowledge extraction and question-answering tasks (Lin et al., 2024). GeologyOracle, built upon the GPT-4o model (OpenAI, 2025) and trained on Earth Science data, can analyze geologic datasets, suggest new geoscience hypotheses, explain Earth-Science concepts, and interpret geosites (Baucon and de Carvalho, 2024). Another project, GeoProspect, introduced one of the first geology-focused LLMs with continual learning, built on a high-quality geological text corpus and evaluated with a tailored benchmark (GeoEval) to ensure it can reason about stratigraphy and mineral exploration scenarios (Wu et al., 2025). Generally speaking, foundation geoscience models can address problems in multimodal and multidimensional datasets with numerical, text, audio, and video inputs Hadid et al. (2024). These specialized LLMs show important progress toward improving the use of AI in geoscience study and application. They provide domain-specific knowledge and skills that general-purpose LLMs could lack, therefore transforming the way geoscientists engage with data, literature, and challenging analytical tasks. The primary advantages of utilizing FMs in geoscience include their scalability (handling large datasets, extensive model training, and high computational demands with ease), generalizability (achieving high performance on novel tasks beyond the initial training scope), dynamic capabilities (such as convolutional operations and support for multiple data modalities), and adaptability (Zhang et al., 2023).

Vision Language Models (VLMs) combine visual and textual data processing capabilities, enabling them to interpret and generate content that integrates both modalities. In geosciences, VLMs have been applied to tasks such as image classification, object detection, and scene understanding, which is also valuable in geoscience contexts. CGG Services has patented a geologic interpretation method (Song, 2024) based on VLMs. Their approach receives verbal and written descriptions of geological features from users, converts them into interpretable input data using a large language model, configures a pretrained VLM based on this data and geological images from other subsurfaces, and delineates geological features in seismic images. This method bridges natural language descriptions with visual geological interpretation. Building on this integration of visual and linguistic modalities, Guo et al. (2025) demonstrated the effectiveness of adapting visual foundation models (VFMs) for geophysical data analysis. By fine tuning pretrained vision encoders on domain-specific datasets such as seismic and lunar imagery, their approach achieved notable improvements in tasks like facies classification and crater detection. This cross-domain adaptation highlights the potential of VLM-based frameworks to leverage pretrained visual models for efficient, accurate interpretation of complex geoscientific imagery.

To further enhance the performance of LLMs in geoscience applications, domain adaptation techniques have been employed to make them more effective. Fan et al. (2025) present a comprehensive survey examining four key adaptation methodologies: prompt engineering, retrieval-augmented generation (RAG), domain-adaptive pretraining, and fine tuning. Their findings demonstrate that domain-adapted LLMs substantially improve reasoning accuracy, automation, and interpretability, though challenges remain in data scarcity, validation, and explainability.

LLM prompts are used to customize the input queries to better align with the model's training and capabilities, thereby enhancing output relevance and accuracy. Task-specific instructional prompts could improve LLM problem-solving accuracy, e.g., in the geotechnical engineering domain (Chen et al., 2024; Fan et al., 2025). Li and Shi (2025) used prompt engineering to enhance generation of realistic geological cross-sections from borehole data and showed that domain-contextual examples can guide the LLM to produce more accurate and relevant outputs. In their approach, a GPT-based model was given a

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

few domain-specific example prompts (few-shot learning) and could then output a 2D cross-sectional sketch consistent with limited borehole data (Li and Shi, 2025). Crucially, they incorporated domain knowledge via curated examples and used self-consistency checks to reduce reasoning errors. This illustrates how providing geological context can steer an LLM to produce technically meaningful visualizations.

Knowledge-graph-augmented prompting represents a notably effective strategy for improving model reliability. By coupling LLMs with structured ontologies describing geotechnical entities – such as geological formations, soil classifications, and loading conditions – these methods embed factual context directly into the reasoning process. In a recent study, Xu et al. (2024) showed that incorporating ontology-based graph information markedly increased the predictive accuracy of LLMs for identifying adverse geological conditions in tunnel construction. While prompt engineering has emerged as a lightweight adaptation technique, RAG has proven as a more comprehensive method for enhancing LLM performance particularly valuable for integrating legacy geological data. For example, Ma et al. (2025) developed a domain-specific RAG pipeline for the Qin-Hang metallogenic belt that leveraged existing geological knowledge bases. Recent implementations combine vector search and generative AI to extract and structure relevant information from unstructured legacy data, utilizing semantic embedding models and large language models to generate accurate, context-aware responses with source traceability (Hungund et al., 2025).

The Geological Everything Model 3D (GEM) (Dou et al., 2025) represents a significant advance as a unified generative architecture that reformulates multiple geological tasks – including fault and horizon delineation, relative geological time estimation, and geobody segmentation – as prompt-conditioned inference. GEM employs a two-stage strategy with self-supervised pretraining on over 500 unlabeled field seismic volumes, followed by fine tuning with heterogeneous prompt-label pairs. This enables zero-shot generalization to unseen tasks and geological settings.

Davis and colleagues developed Geo-LM (Davis, 2025; Cleverley, 2025), a system that leverages Llama 4 (Meta Platforms, 2025) to process geological reports and convert them into 3D geological models. The workflow involves three key steps: (1) document understanding through OCR and image interpretation, (2) consolidation of geological information into a structured "geology DSL" (Domain Specific Language) encoding lithology, structural interpretations, cross-cutting relations, and temporal ordering, and (3) parsing the DSL for input to GemPy (de la Varga et al., 2024) for 3D model generation. GemPy (de la Varga et al., 2024) stands as a prominent open-source Python library for generating 3D structural geological models using implicit modeling approaches. The library employs universal cokriging interpolation supported by advanced mathematical libraries and is capable of constructing complex models including multiple conformal layers, unconformities, magmatic bodies, and fault networks. GemPy's architecture provides efficient probabilistic modeling and uncertainty quantification. Combined with GemPy, the Geo-LM project represents the first demonstration of using LLM APIs to extract and structure information from lengthy geological documents for automated 3D modeling.

In contrast to prompt engeneering or RAG, which influence model behavior during inference, domain-adaptive pretraining functions within the learning stage of model development. As an unsupervised approach, it strengthens an LLM's domain comprehension by extending its pretraining on specialized, domain-specific text corpora (Gururangan et al., 2020). For geological applications, this corpora – text knowledge base – may include academic papers, technical reports, field guides, and

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

other relevant literature. By exposing the model to this specialized vocabulary and context during pretraining, domain-adaptive pretraining enhances its ability to understand and generate domain-relevant content. For example, Ghorbanfekr et al. (2025) illustrated this effect by applying domain-adaptive pretraining to a BERT-based model trained on geotechnical corpora, which led to enhanced understanding of domain-specific technical contexts.

Unlike other domain adaption techniques, fine tuning modifies the model's parameters directly by training it on a labeled dataset specific to the target domain or task (Zhang et al., 2025). This supervised approach allows the model to learn precise mappings between inputs and outputs relevant to geological applications. In geoscience, fine tuning has proven effective for adapting pretrained models to domain-specific linguistic and contextual patterns. The work of Ghorbanfekr et al. (2025) mentioned earlier in the domain-adaptive pretraining context, also demonstrated that application of fine tuning to a BERT-based architecture on geotechnical engineering texts, helped to achieve improved accuracy in understanding geotechnical terminology and contextual relationships within subsurface descriptions. Similarly, Li et al. (2025b) applied a fine-tuned BERT-BiLSTM-CRF framework to geological disaster reports, substantially enhancing named entity recognition performance for key geoscientific terms and event attributes. Large-scale implementations such as K2 (Deng et al., 2024) and GeoProspect (Wu et al., 2025) extend this approach, combining instruction-tuning and supervised fine tuning to align general-purpose foundation models with geological reasoning, terminology, and task requirements, thereby enabling more accurate and context-aware geoscientific analysis.

Overall, these domain-tuned models highlight a trend toward bridging the gap between general LLM capabilities and the nuanced vocabulary and knowledge of geology. They have primarily been applied to tasks like semantic text analysis, question answering, and educational assistance. However, none of these specialized LLMs have yet tackled the geometric or spatial modeling of subsurface structures without relying on external tools.

3 Research Gaps

Despite substantial progress across conventional, machine learning, and generative AI approaches, current approaches still struggle to connect qualitative geological reasoning with automated, geometry-based modeling. Existing methods depend on structured data or predefined workflows, offering limited support for natural language interaction or iterative refinement. This gap highlights the need for a framework that can directly translate geological descriptions into validated, editable cross-section geometries through intuitive, language-driven interaction.

These persistent shortcomings underscore several critical research gaps that the present work aims to address:

1. AI-driven Cross-section Generation: While prior works have explored ML (Lyu et al., 2024; Hillier et al., 2023) or generative AI (Lyu et al., 2024; Ran et al., 2022; Li and Shi, 2025) for geological cross-section generation, they are all data-based, relying on structured inputs like borehole data or seismic images. No existing system directly generates cross-sections from unstructured natural language descriptions.

2. Direct Natural Language to Geometry: Existing systems either use LLMs to generate domain-specific languages that require separate modeling tools (Geo-LM Davis (2025)), or they provide conversational interfaces to existing geological

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

databases (e.g., Geolog-IA (Pozo et al., 2025)), but do not directly generate coordinate-based geometries from textual descriptions.

3. Conversational Refinement: While iterative refinement frameworks exist for LLM outputs in general (Madaan et al., 2023; Xie et al., 2024), no system specifically addresses conversational refinement of geological cross-sections with natural language modification commands.

4. Integrated Validation: Topological validation frameworks exist for geological models (Thiele et al., 2016; Parquer et al., 2025), but these are not integrated with LLM-based generation systems that can iteratively correct violations based on geological principles.

5. Self-Contained Modeling: Systems like GeoINR Hillier et al. (2023) and GemPy (de la Varga et al., 2024) provide excellent implicit modeling capabilities, and Geo-LM (Davis, 2025) demonstrates LLM-to-GemPy integration, but no self-contained system exists where the LLM directly generates valid geometric representations without external computational libraries.

6. Bidirectional Natural Language Interaction: Vision language models for geology (Song, 2024) support natural language prompts for interpretation, but do not support the full cycle of generation from natural language descriptions followed by natural language-driven refinement.

GeoSIRR addresses these gaps by introducing a framework that: (1) directly interprets natural language geological descriptions and generates structured coordinate-based cross-section geometries; (2) supports conversational, iterative refinement through natural language commands; (3) Integrates topological and geological validation within the generation process; (4) operates without requiring external modeling libraries or tools for the core generation functionality (5) Enables bidirectional interaction between qualitative geological reasoning and quantitative geometric modeling.

4 Methodology

The proposed GeoSIRR system introduces a dual-stage framework for the automated interpretation, reconstruction and refinement of geological cross-sections using large language models (LLMs). The methodology integrates natural language understanding, geometric reasoning, and rule-based validation within an interactive environment, allowing users to describe geological structures textually and obtain consistent, coordinate-based visualizations. Conceptually, the modeling process consists of two principal stages: generation and refinement. Figure 1 illustrates the overall workflow of the system for both stages.

4.1 Generation

The generation stage establishes the foundational geometric model from a natural language description. At this stage, GeoSIRR interprets the user's textual input – typically a geological description such as "*a west-dipping thrust fault cutting through three*

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

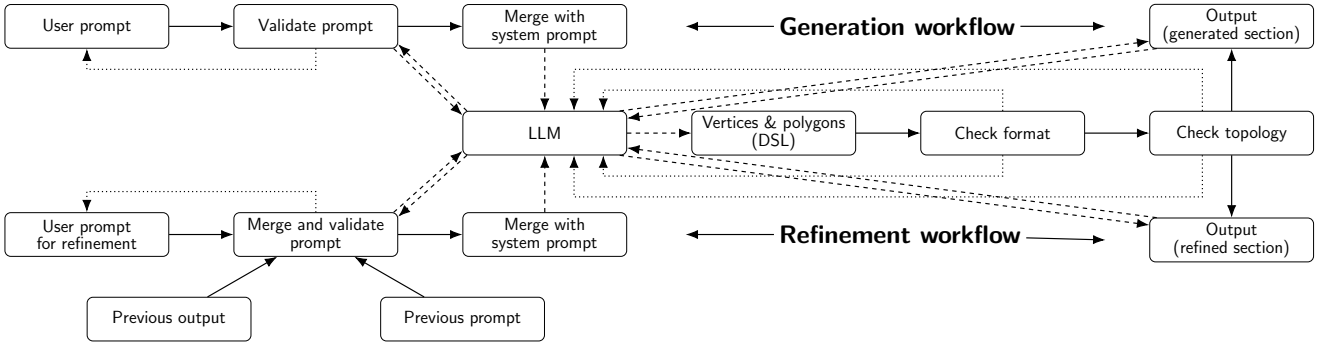

Figure 1. Workflows for the generation (upper) and refinement (lower) stages of the proposed geological cross-section modeling system. In the generation stage, a natural language description is processed by the LLM to produce a structured geometric definition, which is then validated. In the refinement stage, user modification requests are interpreted by the LLM to update the existing geometry while preserving unchanged features, followed by validation of the revised model.

sedimentary layers" – and translates it into a structured representation composed of vertices, coordinates, and polygons that define the spatial relationships between geological bodies.

The LLM acts as an interpretive agent that performs a semantic transformation from descriptive geological language into a consistent numerical and topological definition. It references a set of predefined modeling conventions that describe how geological entities – such as faults, folds, intrusions, or sedimentary layers – are represented as continuous closed polygons defined by sequential vertices (x, z). Each vertex denotes a point in a two-dimensional cross-sectional space, forming part of the boundary of one or more geological bodies. The resulting configuration is a connected, gap-free assembly of polygons that together form a continuous rectangular cross-section.

Prior to generation, an auxiliary "clarification agent" validates the sufficiency and clarity of the textual description. This agent identifies whether essential information – such as the number of layers, structure type, or key dimensions – is present, and when necessary, suggests reasonable defaults. This step ensures that even incomplete or ambiguous geological descriptions can be converted into workable models without extensive user intervention.

Once validated, the generation model constructs the cross-section iteratively. Each iteration refines the geometric consistency of the output while preserving geological plausibility and topological continuity. The final result is expressed in a standardized text-based format containing all vertices and polygons, accompanied by automatically rendered visual previews. Notably, LLM does not generate the cross-section images, but the textual description in the prescribed format (see Section 4.4). Visualization is done independently using Python.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

4.2 Refinement

The refinement stage enables interactive modification of previously generated cross-sections through natural language commands. Instead of manually editing geometric entities, users can request conceptual alterations, for example "*add a dyke cutting across all layers*" or "*increase the dip of the fault to 60°*".

Upon receiving a refinement prompt, the system reloads the existing geometric definition and instructs the language model to regenerate an updated version while preserving all unchanged features. The model analyzes the existing topology, interprets the modification request, and produces a revised definition consistent with both the original geometry and the new geological instruction. All changes are verified for structural integrity through automated validation routines that check format correctness, polygon connectivity, and geological continuity.

This interactive refinement cycle allows users to iteratively evolve a geological concept into a realistic and consistent cross-section, maintaining geological plausibility throughout. The system thereby bridges the gap between qualitative geological reasoning and quantitative modeling, translating conceptual modifications into structured geometric updates.

4.3 Validation and Representation

Both generation and refinement stages employ internal validation mechanisms. These include checks for format correctness (unique and ordered vertices, properly closed polygons) and topological consistency (no overlaps, no internal gaps, rectangular shape of the cross-section). The output is further validated against geological principles. For instance, normal faults must have downward-moving hanging walls or intrusions must cut through pre-existing layers.

The final representation of the cross-section is stored as a human-readable and machine-parsable text definition. It can be rendered into 2D or 3D visualizations and exported for use in external modeling systems. This standardized format also facilitates downstream numerical simulations or geological interpretation workflows.

4.4 Domain-Specific Language (DSL) for Cross-Section Definition

A core component of the proposed framework is a domain-specific language (DSL) designed for the structured definition of geological cross-sections. This DSL serves as the intermediate representation between natural language input and the system's internal geometric model. It provides a consistent syntax for expressing vertices, polygons, and geological bodies in a human-readable yet machine-parsable format, allowing seamless conversion between descriptive geological text and coordinate-based geometry.

4.4.1 Vertices and Polygons

Each geological cross-section is represented as a regular rectangular domain composed of one or more polygons, where each polygon corresponds to a geological body such as a layer, fault block, or intrusion. The geometry of each polygon is defined by a set of vertices, each having a unique identifier (ID) and associated (x, z) coordinates on a two-dimensional plane representing horizontal position and vertical elevation, respectively.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Coordinates are expressed in kilometers to reflect geological scale. The x axis denotes the horizontal distance along the cross-section, and the z axis denotes elevation relative to mean sea level (negative downwards). Elevation here is a better choice compared to the depth (positive downwards) which would be more confusing for the LLM because it can be treated as variable along x (due to changing topographic heights). Thus, $z = 0$ corresponds to the sea level, while negative z values correspond to deeper parts.

We deliberately impose a regular rectangular shape for any generated cross-section, because any topography variations, if required, can be easily incorporated by the LLM itself by inclusion of an "air" unit, as shown in Example 2 later. The main benefit of it is a much more simple and fast way to check the topological correctness of the generated set of polygons, as explained in Section 4.4.3.

An individual vertex record follows the general format:

```
1: <ID> <x> <z>
```

where `<ID>` is a unique integer (starting at 0), and `<x>` and `<z>` are floating-point coordinates. For example, the following lines define four vertices:

```
1: 0 0.0 0.0
2: 1 5.0 0.0
3: 2 5.0 -5.0
4: 3 0.0 -5.0
```

Here, vertex 0 is located at the top-left corner of the cross-section, vertex 1 is positioned 5 km to the right at the top-right corner, vertex 2 is at the bottom-right corner (5 km right, 5 km depth), and vertex 3 is at the bottom-left corner (0 km horizontal, 5 km depth).

Each polygonal geological body is defined by a name and a sequence of vertex IDs listed in a certain order. The polygon's boundary is implicitly closed, meaning the final vertex connects back to the first one. The polygon definition takes the following form:

```
1: <name> <v0> <v1> <v2> ... <vn>
```

where `<name>` is a unique alphanumeric identifier for the geological body beginning with a letter (e.g., "Layer1", "UnitA"), and `<v0> ... <vn>` are the ordered vertex IDs that delineate the polygon's edges. For example, the following line

```
1:  layer 0 1 2 3
```

defines a polygon named `layer` composed of four vertices, from 0 to 3, connected sequentially and closed implicitly. Together, the vertex and polygon definitions provide a complete geometric description of the cross-section given in Listing 1.

Listing 1. DSL for the simple single-layer cross-section (Figure 2)

```
1: 0 0.0 0.0
2: 1 5.0 0.0
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026


```
3: 2 5.0 -5.0
4: 3 0.0 -5.0
5: layer 0 1 2 3
```

Figure 2 illustrates an example of a single-layer geological cross-section represented using this DSL format.

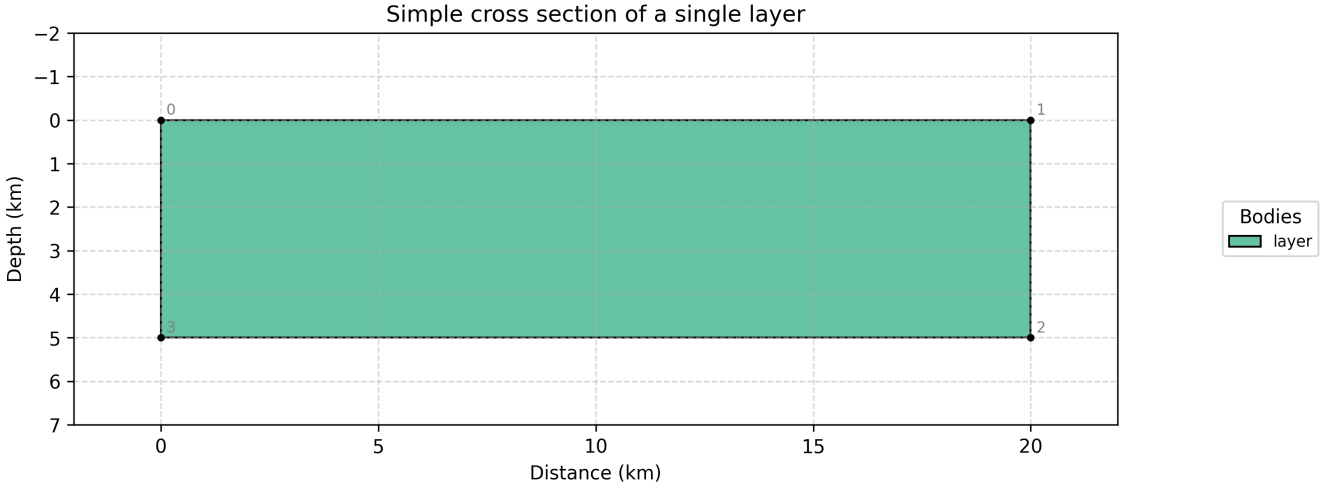

Figure 2. Example of a single-layer geological cross-section represented using the proposed DSL. The layer is defined by four vertices (IDs 0 to 3, depicted with gray) forming a closed polygon.

Vertices can be shared among polygons to ensure geometric continuity between adjacent bodies (e.g., neighboring layers or fault-bounded blocks). This vertex-sharing mechanism avoids redundant coordinates and guarantees that adjacent polygons meet seamlessly without gaps or overlaps – an essential requirement for geological consistency.

Polygons are always arranged to collectively fill the regular rectangular cross-section. The system enforces that all vertices belong to at least one polygon and that polygons connect continuously, preserving the overall rectangular shape of the section.

Listing 2. DSL for the cross-section with two layers (Figure 3)

```
1: 0 0.0 0.0
2: 1 5.0 0.0
3: 2 5.0 -5.0
4: 3 0.0 -5.0
5: 4 0.0 -2.0
6: 5 5.0 -2.0
7: layer1 0 1 5 4
8: layer2 4 5 2 3
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

For instance, Listing 2 defines two layers, `layer1` and `layer2`, sharing vertices 4 and 5 along their common boundary. Figure 3 shows an example of such a geological cross-section with two layers, each defined by its own set of vertices and polygons using the DSL format described above. Note how vertices are shared along the boundary between the two layers to ensure continuity.

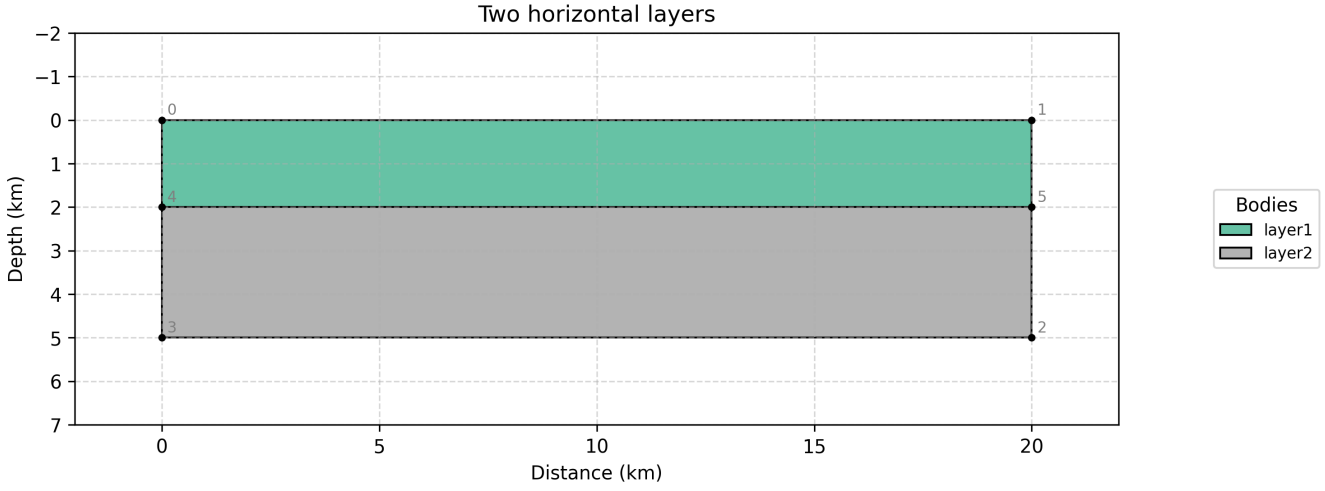

Figure 3. Example of a two-layer geological cross-section represented using the proposed DSL. Each layer is defined by four vertices forming closed polygons.

4.4.2 Handling of Complex Geological Bodies

Geological bodies that cannot be represented by a single contiguous polygon may be expressed as multiple connected components using a consistent base name with part identifiers separated by the ^ symbol (any distinct separator can be used, but it should be consistent across the model).

Listing 3 shows an example of a cross-section with three sedimentary layers on top and one salt body (salt diapir) piercing from below is shown in Figure 4, the middle sedimentary layer (`layer3`) is represented as two separate polygons: `layer3^left` and `layer3^right`. Each part has its own set of vertices but shares a common base name (`layer3`), indicating that they belong to the same geological unit.

In Figure 4, the bottom of the first layer is deformed by the upward intrusion of the salt body, which is represented with additional vertices 17–21 added to capture the complex geometry. The two parts of `layer3` are defined separately to allow for the discontinuity caused by the salt intrusion. This approach allows for flexible representation of complex geological bodies that may be discontinuous or segmented within the cross-section, and extra vertices are used for smooth transitions around such features.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Listing 3. DSL output for the salt diapir (Figure 4)

```
0 0.0 0.0
1 20.0 0.0
2 20.0 -5.0
3 0.0 -5.0
4 0.0 -1.0
5 20.0 -1.0
6 0.0 -2.0
7 20.0 -2.0
8 0.0 -4.0
9 20.0 -4.0
10 7.5 -4.0
11 12.5 -4.0
12 8.5 -2.0
13 11.5 -2.0
14 9.0 -1.7
15 10.0 -1.5
16 11.0 -1.7
17 8.5 -1.0
18 9.0 -0.7
19 10.0 -0.5
20 11.0 -0.7
21 11.5 -1.0
layer1 0 1 5 21 20 19 18 17 4
layer2 4 17 18 19 20 21 5 7 13 16 15 14 12 6
layer3^left 6 12 10 8
layer3^right 13 7 9 11
salt 8 10 12 14 15 16 13 11 9 2 3
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

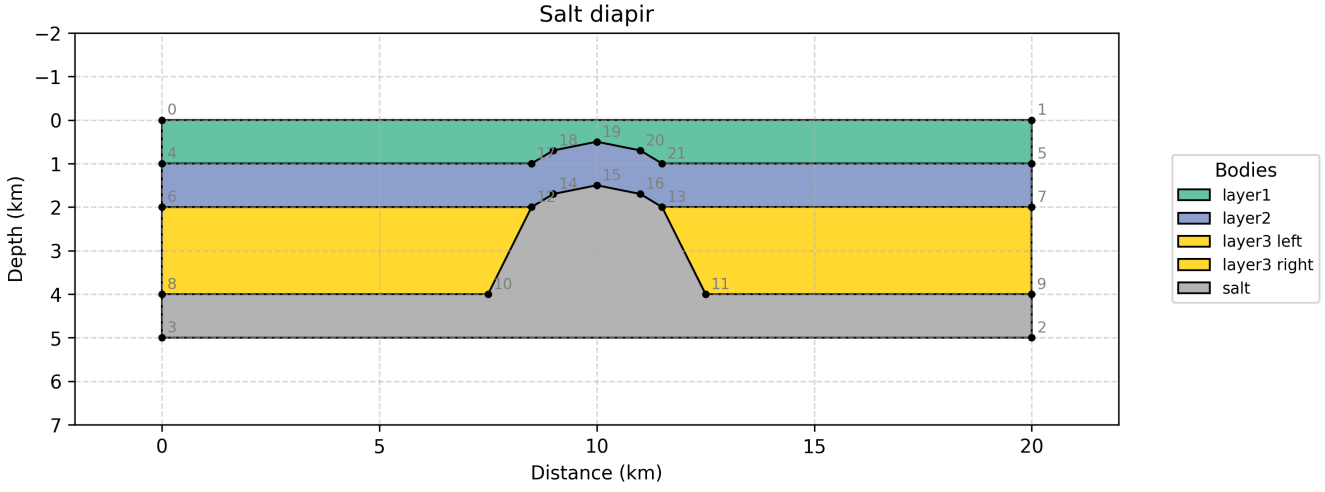

Figure 4. Example of a cross-section with three sedimentary layers and a salt diapir piercing upwards. Note how the middle sedimentary layer is represented as two separate polygons (`layer3^left` and `layer3^right`) to accommodate the intrusion of the salt body.

4.4.3 Topological and Geological Consistency

The DSL inherently supports validation of both geometric and geological constraints. From a geometric perspective, validation ensures proper vertex ordering, absence of internal gaps or overlaps, and continuous polygon connectivity. The basic requirement of a regular rectangular shape of the cross-section domain significantly simplifies these checks. For instance, absence of gaps or overlaps is checked by comparison of the sum of the areas of all the polygons to the area of the domain rectangle.

From a geological standpoint, additional checks can be applied – for instance, ensuring that faults displace layers according to their kinematic sense (normal or reverse) or that intrusions deform pre-existing units.

This structured representation enables the system to reason about geological relationships in a symbolic yet spatially explicit manner, serving as a bridge between natural language geological descriptions and quantitative cross-section models.

5 Results

In order to demonstrate the capabilities of the proposed GeoSIRR system, we present several examples of geological cross-sections generated and refined through natural language descriptions. Each example illustrates different geological scenarios, showcasing the system's ability to interpret complex geological features and produce consistent geometric representations.

For generation of the examples, we utilized GPT-5 (OpenAI, 2025) as the underlying large language model (LLM) due to its advanced natural language understanding and generation capabilities. However, one can use any LLMs with comparable reasoning abilities.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

5.1 Example 1: Listric Normal Fault

Faulted sedimentary layers are common geological features that can be effectively modeled using the proposed system. In this example, we describe a scenario involving a listric normal fault cutting through three sedimentary layers. Listric means that the fault plane curves, typically flattening with depth.

5.1.1 Generation

The original natural language description provided to the system is shown in Listing A1. We use this hierarchical description to prompt the LLM to generate the corresponding cross-section geometry in the most efficient way. The system interprets the geological features, fault characteristics, and stratigraphic layers to produce a structured representation.

The generated cross-section is shown in Figure 5.

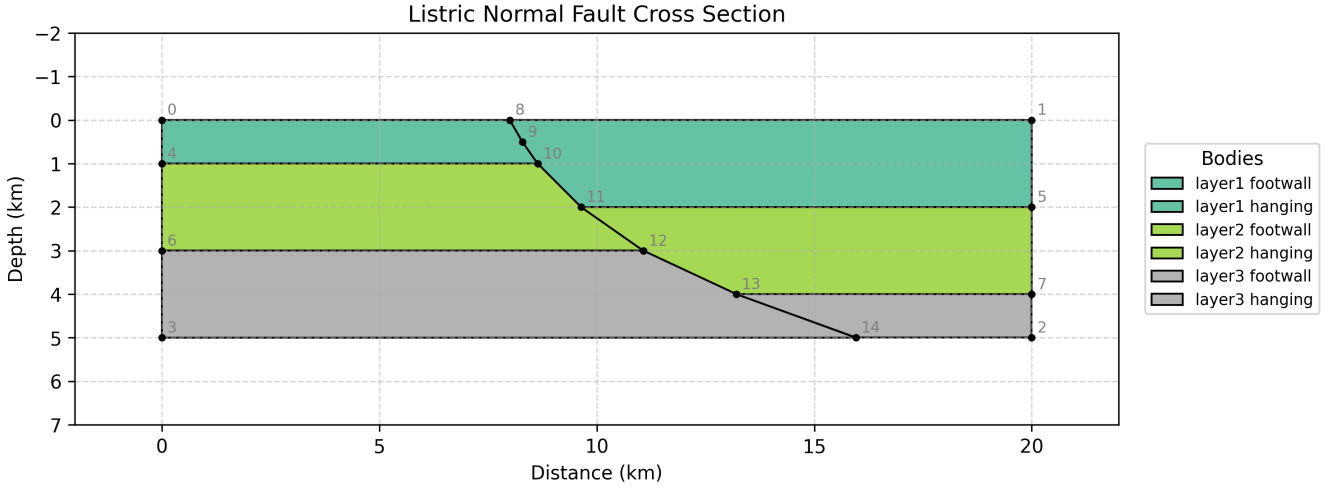

Figure 5. Generated geological cross-section of a listric normal fault cutting through three sedimentary layers. Time for generation with GPT-5 OpenAI (2025): 179.3 seconds.

Note that the system successfully captures the listric geometry of the fault, the displacement of layers, and the varying thicknesses of sedimentary units on either side of the fault. The resulting cross-section in the DSL format described in Section 4.4 is provided in Listing B1.

Note that the comments (lines and blocks starting with #) were generated here by the LLM and help to understand the structure and the reasoning of the LLM, however these comments are optional. Naming of the bodies (e.g., `layer1^footwall`, `layer1^hanging`) follows the conventions outlined in Section 4.4 but are not prescribed – the LLM can choose appropriate names as long as they are unique and consistent.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

5.1.2 Repeatability

One of the concerns with LLM-based generation is the repeatability of results. To assess this, we re-ran the generation process multiple times using the same input description. While minor variations in vertex coordinates and polygon definitions were observed across runs, the overall geological structure and relationships remained consistent. The listric fault geometry, layer thicknesses, and displacement patterns were reliably reproduced, demonstrating the system's robustness in interpreting geological descriptions.

Figure 6 shows the result from one of the repeat runs. The geometry is slightly different in the lower part of the fault, which is constrained by the 20° dip condition at the base in the original description. However, the overall geological features are consistent with the original description and Figure 5, e.g., the starting point of the fault at x=8 km, the hanging wall being downthrown by 1 km, and the layer thicknesses on either side of the fault. The surface dip of the fault is also 60° in the both cases as requested.

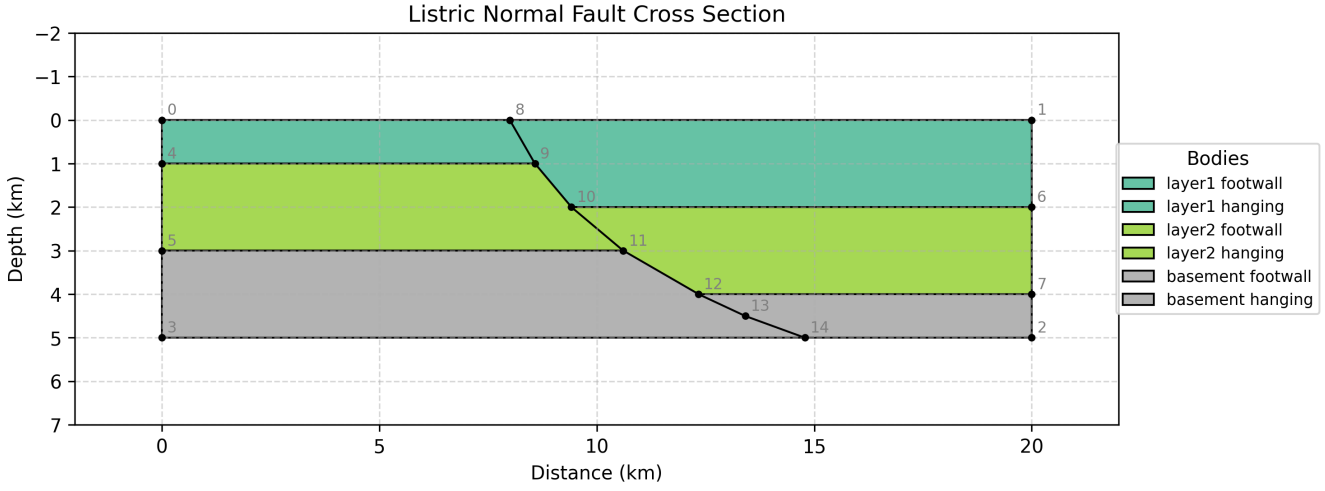

Figure 6. Generated geological cross-section of a listric normal fault cutting through three sedimentary layers (repeat run). Time for generation with GPT-5 OpenAI (2025): 198.1 seconds.

The resulting cross-section in the DSL format from one of the repeat runs (Figure 6) is provided in Listing B2.

Note the variation in the placement of the fault segments (vertices 8 to 14) compared to the first run and the corresponding changes in the polygon definitions, as well as their names. However, the overall geological structure remains topologically correct and consistent with the original description except for the vague constrain on the angle of the fault at depth. Generation time for both runs was also similar, under 200 seconds.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

5.1.3 Refinement

Both Figures 5 and 6 show that the system accurately represents the listric geometry according to the description, with the fault plane curving and flattening with depth, but geologically the geometry is not following the typical fault kinematics, as the hanging wall should expose a rollover fold due to the listric normal faulting.

We used the refinement stage of the system to add this geological feature to the cross-section from Figure 6 by providing the following natural language modification request: *"Add a geologically consistent rollover folding to the hanging wall of the listric fault"*. GeoSIRR produced the updated cross-section shown in Figure 7.

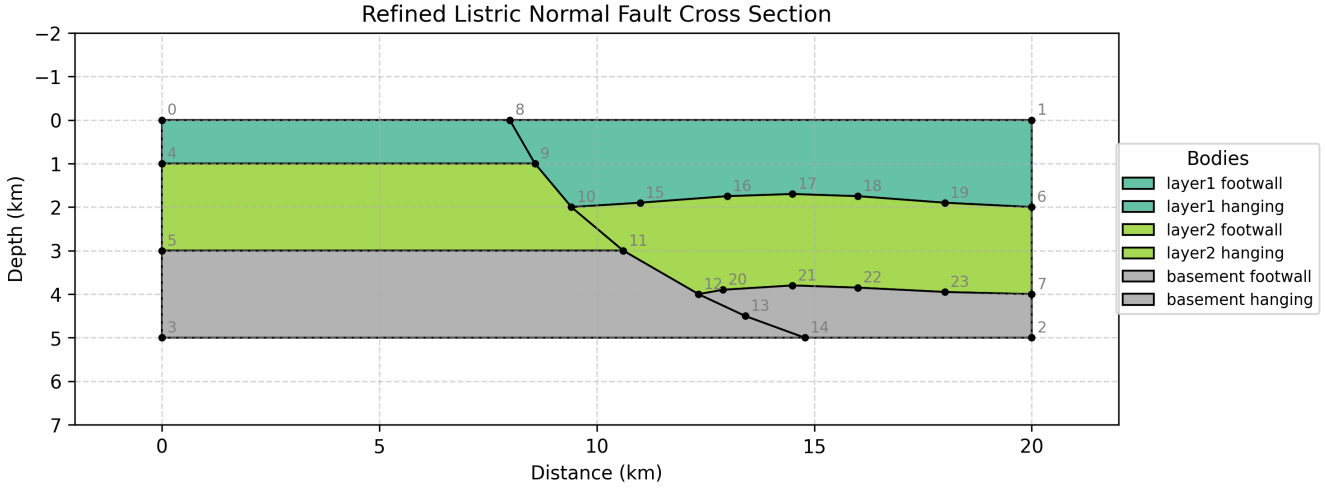

Figure 7. Refined geological cross-section of a listric normal fault (from Figure 6) with added rollover folding in the hanging wall. Time for refinement with GPT-5 OpenAI (2025): 94.8 seconds.

Note that the system successfully incorporated the rollover fold into the hanging wall in a geologically consistent manner, preserving the original fault geometry and layer thicknesses and adding additional vertices to define a smooth fold shape. The time taken for the refinement was significantly shorter (around 100 seconds) than the original generation, demonstrating the efficiency of the refinement process.

In order to illustrate the ability to further refine the model, we designed an additional modification request: *"Add a parallel listric fault with a similar rollover folding starting at x=2 km"*. Resulting cross-section is shown in Figure 8.

Despite the increased complexity of the model, the system effectively added the second listric fault and associated rollover fold while maintaining geological consistency and preserving the original features. The modification request was not overly prescriptive, and didn't specify details about the throw or exact geometry of the new fault, yet the system was able to infer a reasonable representation based on the existing model and geological principles. Note that the remaining features from the previous model – geometry of the first fault, layer thicknesses, and rollover fold – were preserved accurately. The refinement time was longer (around 900 seconds) due to the increased complexity of the geometry.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

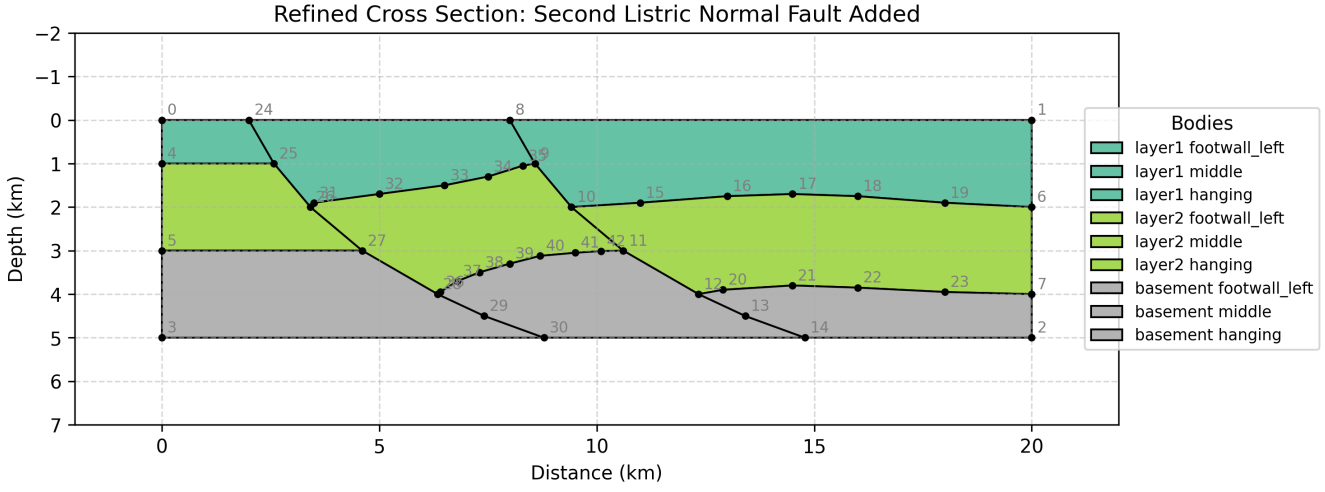

Figure 8. Further refined geological cross-section of a listric normal fault (from Figure 7) with an additional parallel listric fault and rollover folding starting at x=2 km. Time for refinement with GPT-5 (OpenAI, 2025): 892.3 seconds.

5.2 Example 2: Laccolith Dyke Intrusion

Laccoliths are intrusive igneous bodies that cause the overlying strata to dome upwards, often associated with feeder dykes that supply magma from deeper sources. In this example, we describe a scenario in which a dyke intrusion feeds into a laccolith, deforming and uplifting the overlying layers.

5.2.1 Generation

The original natural language description provided to the system is shown in Listing A2. The system interprets the intrusive geometry, including the feeder dyke connection and the doming effect on overlying layers, to produce a structured representation shown in Figure 9. This first-generation example demonstrates handling of a complex scenario involving both vertical (dyke) and horizontal (sill) intrusive elements with smooth tapering geometries and layer deformation. The model successfully bent the upper units over the laccolith while maintaining their constant thickness. Notably, the system automatically added an air layer above the surface where the uplift reaches the surface level, however the original instructions did not specifically mention this. The high concentration of vertices required to represent the smooth tapering mentioned in the prompt resulted in significantly longer generation time (956.6 seconds) compared to the other examples.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

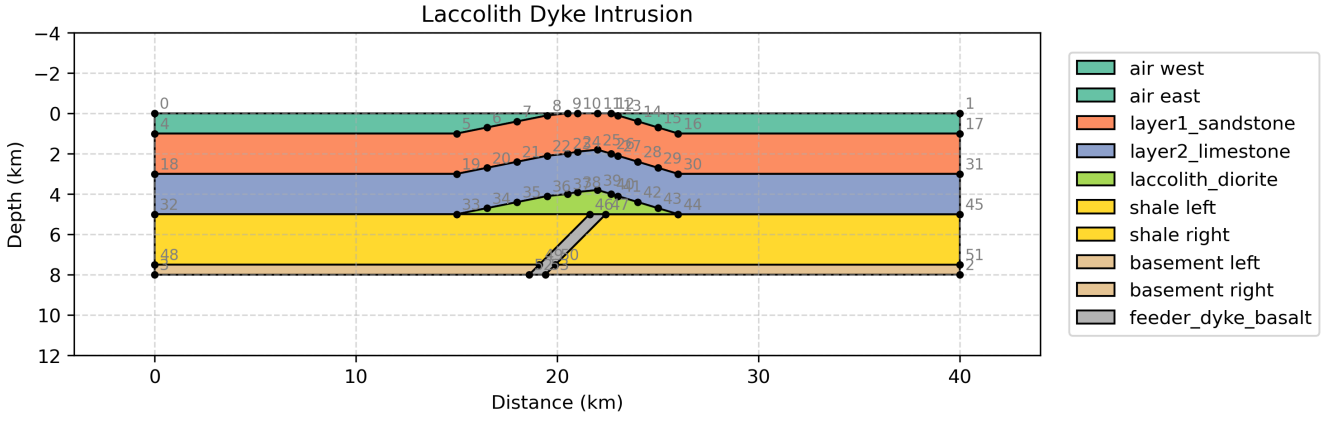

Figure 9. Generated geological cross-section of a laccolith intrusion with feeder dyke. Time for generation with GPT-5 (OpenAI, 2025): 956.6 seconds.

The resulting cross-section in the DSL format described in Section 4.4 is provided in Listing B3.

Note that the system successfully captures the dome geometry of the uplifted layers with smooth tapering of the laccolith from the peak at x=22 km, and the steeply east-dipping feeder dyke (approximately 75°) connecting the deeper source to the laccolith intrusion level. The model preserves the constant thickness of layers 1 and 2 while deforming them over the laccolith dome. The air layer was generated in the area where the uplift brings the sandstone layer to the surface (between x=20.5 km and x=22.6 km), and automatically eroding the top unit, maintaining the rectangular cross-section boundary required by the modeling framework. Note that the sum of the specified layer thicknesses (2 km + 2 km + 2.5 km + 2.5 km = 9 km) exceeds the 8 km depth defined for the section. As a result, the basement layer was reduced to 0.5 km instead of the intended 2.5 km, since the model adjusted the layers to fit within the cross-section limits. This thickness ambiguity likely contributed to the longer generation time (956.6 seconds), as the LLM needed to resolve the conflicting constraints while maintaining geological consistency.

5.2.2 Repeatability

To assess the repeatability of results for this complex geological scenario involving intrusive bodies, we re-ran the generation process twice using the same input description. The system consistently reproduced the key geological features: the steeply east-dipping dyke geometry, the lens-shaped laccolith with smooth tapering on both sides, and the upward deflection of overlying layers preserving their thickness.

Figure 10 shows the result from one of the repeat runs. While minor variations in vertex placement and the exact curvature of the domed layers and laccolith edges were observed, the overall geological structure remained consistent with the original description.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

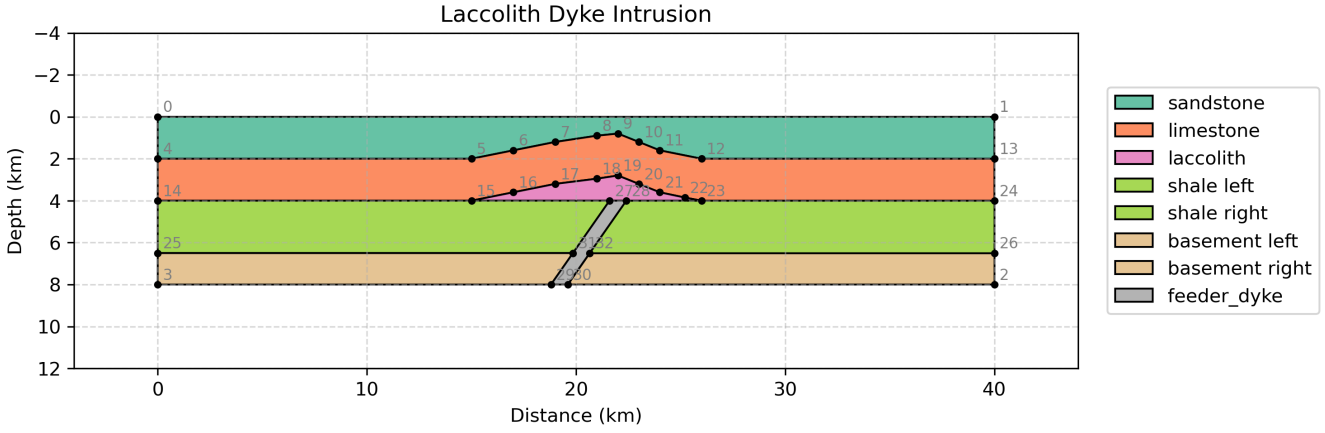

Figure 10. Generated geological cross-section of a laccolith intrusion with feeder dyke (repeat run). Time for generation with GPT-5 (OpenAI, 2025): 846.5 seconds.

The key features preserved across runs include: the feeder dyke extending from x=19 km at depth to x=22 km at the intrusion level with approximately 75° east dip, the laccolith centered at x=22 km with gradual western taper (x=15–22 km) and steeper eastern taper (x=22–26 km), the maximum dome uplift of approximately 1.2 km, and the consistent 2 km thickness of the deformed upper layers. The resulting cross-section from the repeat run in the DSL format is provided in Listing B4.

Some differences were observed in the number and distribution of vertices used to represent the smooth curves, resulting in slight variations in the dome and taper geometry. The color assignments also varied between runs due to different number of units, though this does not affect the geological validity of the model. Notably, this repeat run did not include the explicit air layer above the surface that was generated in the first run, resulting in fewer layers in the model. The basement layer also shows a slight difference in thickness, being slightly thinner in this run.

These variations (particularly the inconsistency in generating the air layer and the varying basement thickness) can be attributed to an important aspect of the input description: the total sum of the specified layer thicknesses (2 km + 2 km + 2.5 km + 2.5 km = 9 km) exceeds the 8 km depth extent specified at the beginning of the prompt. This discrepancy creates an ambiguous constraint that the LLM must resolve, leading to different interpretations across runs. In some cases, the model accommodates the mismatch by adding an air layer and adjusting the basement thickness; in other cases, it omits the air layer and redistributes the thickness differently in the basement layer. This thickness ambiguity also contributes to longer generation times in both runs (956.6 and 846.5 seconds), as the LLM must reconcile the conflicting constraints while maintaining geological consistency. Despite the contradictory depth constraints, the fundamental structural relationships (intrusion geometry, layer deformation patterns, and contact relationships) remained geologically consistent to the input description.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

5.3 Example 3: Prograding Delta

Generating cross-sections and laterally correlating units is crucial in oil and gas exploration that allows geoscientists to have a better understanding of the reservoir facies distribution and their subsurface geometries. Prograding deltaic deposits are characterized by their coarse-grained nature, particularly in their delta front, and they have the possibility of getting sealed by impermeable mud units if the progradation is capped by transgression. These deltaic systems show clinoforms and lateral variation in thicknesses of beds depending on the proximity to the source resulting in pinching out units. In this example, a typical delta progrades onto a relatively gentle slope basin.

5.3.1 Generation

The original natural language description provided to the system is shown in Listing A3.

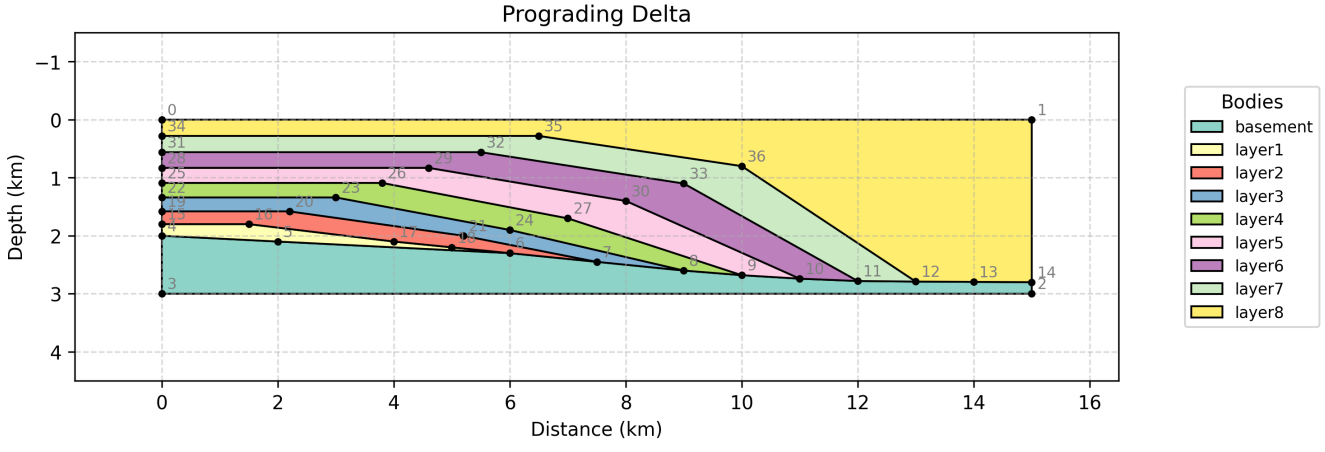

Figure 11. Generated geological cross-section showing progradation of a delta. Time for generation with GPT-5 OpenAI (2025): 751.2 seconds.

The system successfully interprets progradation of various sedimentary packages, and adapt complex terminology such as "pinchout" (Figure 11). This pinch out geometries also followed by downlap reflectors, which is typically visible in seismic surveys. The model also clearly shows a thickening-upward succession demonstrating the ideal and expected stacking pattern.

5.3.2 Interpretation

We also asked the model to make its own geological interpretation to see if it makes sense. The model was asked: *"How was the sea level fluctuation at the time of the layers 1-7 deposition?"*. The answer was as follows: *"Sea-Level Fluctuations Evidence: The clinothem structure, where each layer exhibits a characteristic progradation and pinch-out, indicates a response to relative sea-level fluctuations during deposition. Progradation and Cliniform Geometry: Progradation occurs when sediment supply exceeds accommodation space. The cliniform geometries and rapid eastward thinning of each layer suggest relatively stable or*

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

falling sea levels with periods of relative stasis or stillstands, allowing sediment to prograde. Pinch-Out Patterns: The pinch-out of each layer to the east, coupled with their increasing thicknesses to the west, supports the idea of a regressive sequence in response to a lowering base level, likely reflecting relative sea-level fall or stable low sea levels facilitating progradation. Stacked Clinothems and Progradational Sequences: The succession of clinothems (layers 1 through 7) illustrates a stack of progradational sequences, where each layer represents a depositional episode of deltaic sediment building outward over older deltaic deposits and marine substrates. This pattern is typical of falling or stable sea levels allowing deposition nearer to the shore. In summary, during the deposition of layers 1-7, sea-level stands were likely low or gradually falling, allowing for the progradation of deltaic sediments, with each subsequent layer advancing further eastward over time. Stable sea levels or small fluctuations would have promoted lateral sediment accretion and the development of the cliniform architectures visible in the cross-section."

This geological interpretation is correct and sufficient for a simple delta progradation. The answer is not too simple, and it covers all the possible well-known scenarios. This suggests that the model can also help geoscientists to take their cross section one step further and provide a sufficient interpretation. This can particularly be helpful when complex geometries that can be explained in multiple ways.

5.3.3 Limitations

While GeoSIRR demonstrates the feasibility of generating and refining geological cross-sections directly from natural language descriptions, several limitations should be acknowledged. First, the quality and consistency of the generated models depend strongly on the clarity and completeness of the input description. Ambiguous or internally inconsistent constraints (e.g., conflicting layer thicknesses and section depth) can lead to increased generation time and variability in geometric outcomes, as illustrated in the intrusive body examples. Second, although the framework enforces topological continuity and basic geological rules, the validation mechanisms are rule-based and do not replace expert geological judgment; complex kinematic histories, regional context, or process-based constraints are not explicitly modeled. Third, the current implementation focuses on two-dimensional cross-sections represented by polygonal geometries, and does not directly address three-dimensional modeling, uncertainty quantification, or stochastic ensemble generation. Finally, generation performance and repeatability are influenced by the underlying LLM and its configuration, which may change over time and across model providers. These limitations highlight directions for future development, including tighter constraint handling, expanded geological validation, and extension of the approach to three-dimensional and probabilistic modeling frameworks.

6 Conclusions

While there have been important strides in applying AI and especially LLMs to geoscience, prior works have not addressed the direct generation of geological cross-sections from descriptive language. Current methods either rely on structured inputs (images, borehole tables, etc.) or focus on textual Q&A and coding assistance. No existing framework offers the ability to interpret an arbitrary geological description and produce a consistent, coordinate-based cross-section on the fly, nor to iteratively refine

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

such a model through conversational instructions. This is the critical gap that our proposed GeoSIRR system fills. By integrating an LLM's natural language understanding with a geometry engine and rule-based geological constraints, we enable for the first time a dialogue-driven approach to subsurface structural modeling. In the context of the related work, our contribution can be seen as a novel synthesis: it builds on the trend of domain-specialized LLMs and the early demonstrations of LLM-driven geology tasks, and pushes into new territory by generating a tangible structural representation of the cross-section directly from language. The examples presented illustrate the GeoSIRR's ability to handle complex geological scenarios, including fault geometries and stratigraphic variations caused by intrusive bodies of units, while maintaining geological plausibility and topological consistency. Repeatability tests confirm the robustness of the approach, with consistent geological structures reproduced across multiple runs despite minor geometric variations. Ambiguity in the input descriptions (e.g., conflicting layer thicknesses and section depth) can lead to different interpretations, highlighting the importance of clear and precise geological narratives for optimal results. The refinement capabilities further enhance the user experience, allowing for iterative model evolution based on conceptual modifications, even in complex scenarios. This approach complements existing techniques – offering a more intuitive, high-level mode of interaction – and stands to significantly streamline the workflow of constructing and iterating on geological models. Ultimately, it bridges the long-standing divide between qualitative geological narratives and quantitative model building, positioning the present work as a first-of-its-kind system at the intersection of LLM-driven interpretation and subsurface modeling. By embedding geological reasoning, geometric formalism, and linguistic understanding into a unified framework, GeoSIRR allows efficient, intuitive, and geologically consistent cross-section construction suitable for educational, exploratory, and professional applications.

Code availability. The source code for the GeoSIRR and data can be freely downloaded from https://github.com/CPG-KFUPM/GeoSIRR (last access: 30 December 2025) or https://doi.org/10.5281/zenodo.18097054 (Anikiev and Mosquera, 2025).

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Appendix A: Input Prompts

This appendix contains the complete natural language prompts provided to the system for the examples presented in Section 5.

A1 Example 1: Listric Normal Fault

Listing A1. Input prompt for the Listric Normal Fault example

```
# Listric Normal Fault Cross Section

## Section Overview
A vertical cross-section showing a **listric normal fault** in an extensional tectonic setting.

## Section Extent
* **Horizontal:** 0 km to 20 km
* **Vertical:** 0 km (surface) to 5 km (depth)

## Geological Features

### Fault F1 (Normal Fault)
- **Location:** Surface trace at x = 8 km
- **Dip:** 60^\circ to the east (dipping RIGHT/eastward)
- **Displacement:** 1 km vertical throw
- **Type:** Normal fault - extensional
- **Curvature** Curving eastwards, flattening to a 20^\circ dip at the depth of 5 km. Curvature is approximated with 6 short straight segments.
- **Motion:** Hanging wall (EAST side - above the eastward-dipping plane) moves DOWN

### Structural Blocks
1. **Western Block (Footwall)** - 0-8 km
   - Relatively uplifted (higher stratigraphic position)

2. **Eastern Block (Hanging Wall)** - 8-20 km
   - Downthrown by 1 km (layers at deeper depths)

## Stratigraphic Layers

### Layer 1 (Top - Youngest)
- **Lithology:** Sandstone and shale
- **Thickness:** 1 km in the footwall, thickening to 2 km in the hanging wall (due to syn-tectonic deposition).

### Layer 2 (Middle)
- **Lithology:** Limestone
- **Thickness:** 2 km

### Layer 3 (Bottom - Oldest)
- **Lithology:** Basement rocks
- **Thickness:** 2 km in the footwall, thinning to 1 km in the hanging wall to accommodate fault displacement and section base.
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

A2 Example 2: Laccolith Dyke Intrusion

Listing A2. Input prompt for the Laccolith Dyke example

```
# Laccolith Intrusion Cross Section

## Section Extent
* **Horizontal:** 0 km to 40 km
* **Vertical:** 0 km (surface) to 8 km (depth)

## Geological Features

### Feeder Dyke
- **Width:** 0.8 km
- **Dip:** 75^ east
- **Location:** Bottom at x=19 km (depth 8 km), top at x=22 km (depth 5 km)
- **Lithology:** Basalt

### Laccolith Sill
- **Type:** Lens-shaped intrusion at top of shale layer (5 km depth)
- **Peak:** x=22 km, thickness 1.2 km
- **West side:** Gradual taper from x=15 km to x=22 km
- **East side:** Steeper taper from x=22 km to x=26 km
- **Edges:** Taper smoothly to zero thickness
- **Top:** Smooth dome shape
- **Lithology:** Diorite

### Layer Deformation
- Layers 1 and 2 bend upward over laccolith, maintaining constant 2 km thickness
- Maximum uplift ~1.2 km at x=22 km
- Layers return to flat beyond x=15 km (west) and x=26 km (east)

## Stratigraphic Layers
- **Layer 1 (Sandstone):** 2 km thick, deformed over laccolith
- **Layer 2 (Limestone):** 2 km thick, deformed over laccolith
- **Layer 3 (Shale):** 2.5 km thick, laccolith intrudes at top
- **Layer 4 (Basement):** 2.5 km thick, flat
```

A3 Example 3: Prograding Delta

Listing A3. Input prompt for the Prograding Delta example

```
Draw a w-E cross-section showing an overall progradation of a delta. The basement is 1 km thick in the W and 0.2 km in the E.
W-E is 15 km in length.
At 2 km the basement shows thinning towards W. Above the basement, there are 8 layers. In the west, the one at the bottom is 0.2 km in thickness and as you go up in the section the thicknesses of the layers gradually increase to 0.5 km. These layers gradually pinchout as you go towards E and overall thicknes reach to 2 km. The section is progradational, and units show clear cliniforms.
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Appendix B: DSL Output Listings

This appendix contains the complete DSL output listings generated by the system for the examples presented in Section 5.

B1 Example 1: Listric Normal Fault

Listing B1. Initial generation: DSL output for the Listric Normal Fault (Figure 5)

```
0 0.0 0.0 # top left corner of the cross section (surface)
1 20.0 0.0 # top right corner of the cross section (surface)
2 20.0 -5.0 # bottom right corner of the cross section
3 0.0 -5.0 # bottom left corner of the cross section

4 0.0 -1.0 # footwall: base of Layer 1 / top of Layer 2
5 20.0 -2.0 # hanging wall: base of Layer 1 / top of Layer 2
6 0.0 -3.0 # footwall: base of Layer 2 / top of Layer 3
7 20.0 -4.0 # hanging wall: base of Layer 2 / top of Layer 3

## Fault F1 (normal, listric, east-dipping): 6 straight segments approximating curvature; vertical throw \approx 1 km
8 8.0 0.0 # F1 surface trace (x=8 km)
9 8.289 -0.5 # F1 segment (steep \approx 60^\circ)
10 8.639 -1.0 # F1 at z=-1 km (Layer1 FW/HW boundary tie)
11 9.639 -2.0 # F1 at z=-2 km (Layer1^right base / Layer2^right top)
12 11.067 -3.0 # F1 at z=-3 km (Layer2 FW/HW boundary tie)
13 13.212 -4.0 # F1 at z=-4 km (Layer2^right base / Layer3^right top)
14 15.959 -5.0 # F1 at z=-5 km (flattened \approx 20^\circ at base)

## Bodies
layer1^footwall 0 8 9 10 4 # Layer 1 on footwall: 1 km thick (0 to -1)
layer1^hanging 8 1 5 11 10 9 # Layer 1 on hanging wall: thickens to 2 km (0 to -2)

layer2^footwall 4 10 11 12 6 # Layer 2 on footwall: 2 km thick (-1 to -3)
layer2^hanging 11 5 7 13 12 # Layer 2 on hanging wall: 2 km thick (-2 to -4)

layer3^footwall 6 12 13 14 3 # Layer 3 on footwall: 2 km thick (-3 to -5)
layer3^hanging 13 7 2 14 # Layer 3 on hanging wall: thins to 1 km (-4 to -5)
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Listing B2. Repeat run: DSL output for the Listric Normal Fault (Figure 6)

```
0 0.0 0.0 # top left corner of cross section (surface)
1 20.0 0.0 # top right corner (surface)
2 20.0 -5.0 # bottom right corner
3 0.0 -5.0 # bottom left corner

4 0.0 -1.0 # footwall: top of Layer 2 (base of Layer 1), 1 km thick L1 on footwall
5 0.0 -3.0 # footwall: top of Basement (base of Layer 2)

6 20.0 -2.0 # hanging wall: top of Layer 2 (base of Layer 1), 2 km thick L1 on hanging wall
7 20.0 -4.0 # hanging wall: top of Basement (base of Layer 2)

## Fault F1 (listric normal fault, dips east; 6 straight segments approximating curvature; 1 km vertical throw)
8 8.0 0.0 # F1 surface trace (x=8 km), dip \approx 60^\circ here
9 8.577 -1.0 # F1 at z=-1 km (segment \approx 60^\circ)
10 9.416 -2.0 # F1 at z=-2 km (segment \approx 50^\circ)
11 10.608 -3.0 # F1 at z=-3 km (segment \approx 40^\circ)
12 12.340 -4.0 # F1 at z=-4 km (segment \approx 30^\circ)
13 13.412 -4.5 # F1 at z=-4.5 km (segment \approx 25^\circ)
14 14.785 -5.0 # F1 at z=-5 km (segment \approx 20^\circ, fault root at section base)

## Bodies (three layers split by the east-dipping normal fault F1; hanging wall is downthrown by 1 km)
layer1^footwall 0 8 9 4 # Layer 1 (top) on footwall: surface to z=-1, 0-8.577 km until F1
layer1^hanging 8 1 6 10 9 # Layer 1 (top) on hanging wall: surface to z=-2, 8-20 km; includes F1 segment 10-9

layer2^footwall 4 9 10 11 5 # Layer 2 (middle) on footwall: z=-1 to z=-3 up to F1 (uses F1 9-10-11)
layer2^hanging 10 6 7 12 11 # Layer 2 (middle) on hanging wall: z=-2 to z=-4 (uses F1 12-11)

basement^footwall 5 11 12 13 14 3 # Basement on footwall: z=-3 to -5 (uses F1 11-12-13-14)
basement^hanging 12 7 2 14 13 # Basement on hanging wall: z=-4 to -5 (uses F1 14-13-12)
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

B2 Example 2: Laccolith Dyke Intrusion

Listing B3. Initial generation: DSL output for the Laccolith Dyke (Figure 9)

```
0 0.0 0.0 # top left corner of cross section
1 40.0 0.0 # top right corner of cross section
2 40.0 -8.0 # bottom right corner of cross section
3 0.0 -8.0 # bottom left corner of cross section

# Top of Layer 1 (sandstone) - deformed over laccolith; clipped to surface where uplift reaches z=0
4 0.0 -1.0
5 15.0 -1.0
6 16.5 -0.7
7 18.0 -0.4
8 19.5 -0.1
9 20.5 0.0 # uplift reaches surface (west contact)
10 21.0 0.0
11 22.0 0.0
12 22.6667 0.0 # uplift returns below surface east of here
13 23.0 -0.1
14 24.0 -0.4
15 25.0 -0.7
16 26.0 -1.0
17 40.0 -1.0

# Top of Layer 2 (limestone) = base of sandstone; constant 2 km below sandstone top
18 0.0 -3.0
19 15.0 -3.0
20 16.5 -2.7
21 18.0 -2.4
22 19.5 -2.1
23 20.5 -2.0
24 21.0 -1.9
25 22.0 -1.8 # max uplift: 1.2 km (from -3.0 to -1.8) at x=22
26 22.6667 -2.0
27 23.0 -2.1
28 24.0 -2.4
29 25.0 -2.7
30 26.0 -3.0
31 40.0 -3.0

# Base of Layer 2 (equals top of shale outside laccolith; equals laccolith top within 15-26 km)
32 0.0 -5.0
33 15.0 -5.0 # west laccolith edge (pinch-out)
34 16.5 -4.7
35 18.0 -4.4
36 19.5 -4.1
37 20.5 -4.0
38 21.0 -3.9
39 22.0 -3.8 # laccolith peak thickness 1.2 km (top at -3.8)
40 22.6667 -4.0
41 23.0 -4.1
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026


```
42 24.0 -4.4
43 25.0 -4.7
44 26.0 -5.0 # east laccolith edge (pinch-out)
45 40.0 -5.0

# Feeder dyke top (contacts laccolith base at z = -5)
46 21.6 -5.0 # dyke top-left
47 22.4 -5.0 # dyke top-right

# Shale bottom (top of basement), flat at z=-7.5; intersections with dyke sides
48 0.0 -7.5
49 19.1 -7.5 # dyke west wall at shale bottom
50 19.9 -7.5 # dyke east wall at shale bottom
51 40.0 -7.5

# Dyke bottom at section base (z=-8), width 0.8 km
52 18.6 -8.0 # dyke bottom-left
53 19.4 -8.0 # dyke bottom-right

## Bodies
air^west 0 9 8 7 6 5 4 # air above sandstone on the west side (x <= 20.5 km)
air^east 12 1 17 16 15 14 13 # air above sandstone on the east side (x >= 22.6667 km)

layer1_sandstone 4 5 6 7 8 9 10 11 12 13 14 15 16 17 31 30 29 28 27 26 25 24 23 22 21 20 19 18 # sandstone, top clipped at surface where uplift reaches z=0

layer2_limestone 18 19 20 21 22 23 24 25 26 27 28 29 30 31 45 44 43 42 41 40 39 38 37 36 35 34 33 32 # limestone, constant 2 km thick, bent over laccolith

laccolith_diorite 33 34 35 36 37 38 39 40 41 42 43 44 47 46 # laccolith (diorite) sill: top domed (33-44), base along z=-5 including dyke contact (46-47)

shale^left 32 33 46 49 48 # shale west of dyke, from z=-5 to z=-7.5
shale^right 47 44 45 51 50 # shale east of dyke, from z=-5 to z=-7.5

basement^left 48 49 52 3 # basement west of dyke, z=-7.5 to -8.0
basement^right 50 51 2 53 # basement east of dyke, z=-7.5 to -8.0

feeder_dyke_basalt 46 47 50 53 52 49 # 0.8 km wide feeder dyke connecting to laccolith at z=-5 (includes wall points 49,50)
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Listing B4. Repeat run: DSL output for the Laccolith Dyke (Figure 10)

```
0 0.0 0.0 # top left corner of the cross section
1 40.0 0.0 # top right corner of the cross section
2 40.0 -8.0 # bottom right corner of the cross section
3 0.0 -8.0 # bottom left corner of the cross section

# Top-of-L2 (base of Layer 1) - domed over laccolith, returns flat outside [15,26]
4 0.0 -2.0
5 15.0 -2.0
6 17.0 -1.6
7 19.0 -1.2
8 21.0 -0.9
9 22.0 -0.8 # maximum uplift of L2 top (~1.2 km uplift from -2.0 to -0.8) at x=22
10 23.0 -1.2
11 24.0 -1.6
12 26.0 -2.0
13 40.0 -2.0

# Base-of-L2 = top-of-shale or laccolith top within [15,26] (shaped dome), flat at -4 elsewhere
14 0.0 -4.0
15 15.0 -4.0 # west edge of laccolith base on shale top
16 17.0 -3.6
17 19.0 -3.2
18 21.0 -2.95
19 22.0 -2.8 # laccolith top apex (thickness 1.2 km over flat base at -4.0)
20 23.0 -3.2
21 24.0 -3.6
22 25.2 -3.85
23 26.0 -4.0 # east edge of laccolith base on shale top
24 40.0 -4.0

# Shale base (top of basement), flat
25 0.0 -6.5
26 40.0 -6.5

# Feeder dyke top intersection with shale top (aligned at z=-4 for clean topology)
27 21.6 -4.0 # dyke top-left on shale top
28 22.4 -4.0 # dyke top-right on shale top

# Feeder dyke bottom intersections with base of section
29 18.8 -8.0 # dyke bottom-left on base
30 19.6 -8.0 # dyke bottom-right on base

# Feeder dyke intersections with shale base (z=-6.5)
31 19.85 -6.5 # left edge at shale base
32 20.65 -6.5 # right edge at shale base

## Bodies
# Layer 1 (Sandstone): surface to top-of-L2 (surface flat, L2 top domed)
sandstone 0 1 13 12 11 10 9 8 7 6 5 4

```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026


```
# Layer 2 (Limestone): top-of-L2 dome down to base-of-L2 (which equals laccolith top inside [15,26] and flat -4 outside)
limestone 4 5 6 7 8 9 10 11 12 13 24 23 22 21 20 19 18 17 16 15 14

# Laccolith sill (Diorite): flat base at -4 (x=15 to 26, split by dyke top at 27-28), domed top peaking at -2.8 at x=22
laccolith 15 27 28 23 22 21 20 19 18 17 16

# Shale (Layer 3) split into left and right parts by the feeder dyke; both sit between -4 and -6.5
shale^left 14 15 27 31 25
shale^right 28 23 24 26 32

# Basement (Layer 4) split by the feeder dyke; flat between -6.5 and -8.0
basement^left 25 31 29 3
basement^right 32 26 2 30

# Feeder dyke (Basalt), ~0.8 km wide, dipping obliquely; connects basement to shale top
feeder_dyke 27 28 32 30 29 31
```

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Author contributions. DA conducted the research, designed the methodology for generation and refinement of cross-sections, developed the domain-specific language (DSL) for geometric representation, designed and developed the format and topology validation mechanisms, designed the geological example 1, and drafted the original manuscript. JMR developed LLM prompts for geological validation, designed and developed the graphical interface for the framework, designed the geological example 2, and contributed to the writing of the original draft of the manuscript. KA contributed to the conceptualization of the framework and analysis of modeling results from a geological point of view, designed the geological example 3, and revised the manuscript. JB helped with the conceptualization of the overarching research objectives, performed analysis of modeling results from a geological point of view, and revised the manuscript. UBW provided insights into the application of LLMs, revised the manuscript and supervised the research.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors acknowledge the support received from the College of Petroleum Engineering & Geosciences (CPG) of the King Fahd University of Petroleum and Minerals (KFUPM). The authors used AI-based tools (namely, GPT-5 by OpenAI) to assist with refining the language in some parts of this manuscript (apart from the parts related to the results of cross-section generation). The authors reviewed and edited the content as needed and take full responsibility for the integrity of the work.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

References

Abubakar, A., Di, H., Kaul, A., Li, C., Li, Z., Simoes, V., Truelove, L., and Zhao, T. (2022). Deep learning for end-to-end subsurface modeling and interpretation: An example from the Groningen gas field. *The Leading Edge*, 41(4):259–267.

Anikiev, D. and Mosquera, J. E. (2025). GeoSIRR: Geological Section Interpretation, Reconstruction & Refinement (1.0.0). Zenodo (Software). doi:10.5281/zenodo.18097054.

Baucon, A. and de Carvalho, C. N. (2024). Can AI get a degree in Geoscience? Performance analysis of a GPT-based artificial intelligence system trained for Earth Science (GeologyOracle). *Geoheritage*, 16(4).

Boggs, S. (2006). *Basin Analysis, Tectonics, and Sedimentation*, chapter 16, pages 550–581. Prentice Hall, Upper Saddle River, New Jersey, USA, 4 edition.

Bond, C. E. (2015). Uncertainty in structural interpretation: Lessons to be learnt. *Journal of Structural Geology*, 74:185–200.

Bonham-Carter, M. (2016). *Geostatistical Geologic Modeling*, chapter 3.2.2, pages 38–50. Springer, Dordrecht, Netherlands.

Cao, X., Liu, Z., Hu, C., Song, X., Quaye, J. A., and Lu, N. (2024). Three-dimensional geological modelling in Earth Science research: An in-depth review and perspective analysis. *Minerals*, 14(7):686.

Catuneanu, O. (2006). *Principles of Sequence Stratigraphy*. Elsevier.

Caumon, G. (2010). Towards stochastic time-varying geological modeling. *Mathematical Geosciences*, 42(5):555–569.

Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., and Sausse, J. (2009). Surface-based 3d modeling of geological structures. *Mathematical Geosciences*, 41(8):927–945.

Chan, S. and Elsheikh, A. H. (2017). Parametrization and generation of geological models with generative adversarial networks.

Chen, L., Tophel, A., Hettiyadura, U., and Kodikara, J. (2024). An investigation into the utility of large language models in geotechnical education and problem solving. *Geotechnics*, 4(2):470–498.

Chen, X., Wang, H., Zhu, Y., Wang, H., Su, M., Bao, S., and Wu, Y. (2022). Interactive geological data visualization in an immersive environment. *ISPRS International Journal of Geo-Information*, 11(3):176.

Cleverley, P. (2025). Geo-LLM: Using Llama 4 to create 3D geological models from geological reports.

Coiffier, G., Renard, P., and Lefebvre, S. (2020). 3D geological image synthesis from 2D examples using generative adversarial networks. *Frontiers in Water*, 2.

Craigie, N. (2018). *Principles of Elemental Chemostratigraphy: A Practical User Guide*. Springer.

Davis, W. (2025). Williamjsdavis/geo-lm: Geologic models from llama 4 language model + gempy!

de la Varga, M., Schaaf, A., and Wellmann, F. (2024). Gempy 1.0: open-source stochastic geological modeling and inversion.

Deng, C., Zhang, T., He, Z., Chen, Q., Shi, Y., Xu, Y., Fu, L., Zhang, W., Wang, X., Zhou, C., Lin, Z., and He, J. (2024). K2: A foundation language model for geoscience knowledge understanding and utilization. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24. ACM.

Di Federico, G. and Durlofsky, L. J. (2025). Latent diffusion models for parameterization of facies-based geomodels and their use in data assimilation. *Computers & Geosciences*, 194:105755.

Dou, Y., Wu, X., Bangs, N. L., Sethi, H. S., Li, J., Gao, H., and Guo, Z. (2025). Geological Everything Model 3D: A promptable foundation model for unified and zero-shot subsurface understanding.

Dramsch, J. S. (2020). *70 years of machine learning in geoscience in review*, pages 1–55. Elsevier.

Fan, L., Liu, F., and Chen, C. (2025). Domain adaptation of large language models for geotechnical applications.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Fossen, H. (2016). *Structural Geology*. Cambridge University Press, 2 edition.

Fu, Y., Wang, M., Wang, C., Dong, S., Chen, J., Wang, J., Yu, H., Huang, J., Chang, L., and Wang, B. (2025). GeoMinLM: A large language model in geology and mineral survey in Yunnan Province. *Ore Geology Reviews*, 182:106638.

Ghorbanfekr, H., Kerstens, P. J., and Dirix, K. (2025). Classification of geological borehole descriptions using a domain adapted large language model. *Applied Computing and Geosciences*, 25:100229.

Ghyselincks, S., Okhmak, V., Zampini, S., Turkiyyah, G., Keyes, D., and Haber, E. (2025). Synthetic geology – structural geology meets deep learning.

Glagolev, V. A. and Petrishchevsky, A. M. (2010). A simple algorithm for generating cross sections using 3D geological and geophysical data sets (Southeastern Russia). *Russian Journal of Pacific Geology*, 4(3):242–249.

Grose, L., Aillères, L., Laurent, G., Armit, R., Jessell, M., and Caumon, G. (2021). Loopstructural 1.0: Time-aware geological modelling. *Geoscientific Model Development*, 14(6):3915–3937.

Guo, Z., Wu, X., Liang, L., Sheng, H., Chen, N., and Bi, Z. (2025). Cross-domain foundation model adaptation: Pioneering computer vision models for geophysical data analysis. *Journal of Geophysical Research: Machine Learning and Computation*, 2(1).

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., and Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks.

Hadid, A., Chakraborty, T., and Busby, D. (2024). When geoscience meets generative AI and large language models: Foundations, trends, and future challenges. *Expert Systems*, 41(10).

Hassanzadeh, A., Vázquez-Suñé, E., Corbella, M., and Criollo, R. (2022). An automatic geological 3D cross-section generator: Geopropy, an open-source library. *Environmental Modelling & Software*, 149:105309.

Hillier, M., Wellmann, F., de Kemp, E. A., Brodaric, B., Schetselaar, E., and Bédard, K. (2023). Geoinr 1.0: an implicit neural network approach to three-dimensional geological modelling. *Geoscientific Model Development*, 16(23):6987–7012.

Holm, K. O. M. (2024). *GeoGPT: An LLM-Based Geographic Information System*. Msc thesis, NTNU, Trondheim, Norway. Available at https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3155563.

Huang, D., Zuo, R., Wang, J., and Tolosana-Delgado, R. (2025). Spatially constrained variational autoencoder for geochemical data denoising and uncertainty quantification. *Journal of Earth Science*, 36(5):2317–2336.

Hungund, B., Nair, G., Rathore, P. S., and Toms, J. (2025). Unveiling hidden insights: A GenAI-powered RAG framework for legacy well completion reports. In *Fifth EAGE Digitalization Conference & Exhibition*, pages 1–4. European Association of Geoscientists & Engineers.

Jessell, M. W. and Valenta, R. K. (1996). *Structural geophysics: Integrated structural and geophysical modelling*, pages 303–324. Elsevier.

Jones, R. W. and Simmons, M. D. (1999). *Biostratigraphy in Production and Development Geology*, volume 152 of *Geological Society Special Publication*. Geological Society, London.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2019). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554.

Kuckreja, K., Danish, M. S., Naseer, M., Das, A., Khan, S., and Khan, F. S. (2023). GeoChat: Grounded large vision-language model for remote sensing.

Lee, D., Ovanger, O., Eidsvik, J., Aune, E., Skauvold, J., and Hauge, R. (2025). Latent diffusion model for conditional reservoir facies generation. *Computers & Geosciences*, 194:105750.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Li, H. and Shi, C. (2025). Few-shot learning of geological cross-sections from sparse data using large language model. *Geodata and AI*, 2:100010.

Li, S., Qin, Z., Ding, Y., Wang, M., and Hu, B. (2025a). A real-time geological sections generation method for geological 3d models based on per-pixel linked lists. *International Journal of Digital Earth*, 18(1).

Li, Y., Luo, L., Zeng, X., and Han, Z. (2025b). Fine-tuned bert-bilstm-crf approach for named entity recognition in geological disaster texts. *Earth Science Informatics*, 18(2).

Lin, Z., Deng, C., Zhou, L., Zhang, T., Xu, Y., Xu, Y., He, Z., Shi, Y., Dai, B., Song, Y., Zeng, B., Chen, Q., Miao, Y., Xue, B., Wang, S., Fu, L., Zhang, W., He, J., Zhu, Y., Wang, X., and Zhou, C. (2024). Geogalactica: A scientific large language model in geoscience.

Liu, Q. and Ma, J. (2024). Foundation models for geophysics: Reviews and perspectives.

Lopez-Alvis, J., Nguyen, F., Looms, M. C., and Hermans, T. (2022). Geophysical inversion using a variational autoencoder to model an assembled spatial prior uncertainty. *Journal of Geophysical Research: Solid Earth*, 127(3).

Lyu, B., Wang, Y., and Shi, C. (2024). Multi-scale generative adversarial networks (gan) for generation of three-dimensional subsurface geological models from limited boreholes and prior geological knowledge. *Computers and Geotechnics*, 170:106336.

Ma, J., Zhou, Y., He, L., Zhang, Q., Bilal, M. A., and Zhang, Y. (2025). Enhancing geological knowledge engineering with retrieval-augmented generation: A case study of the Qin-Hang metallogenic belt. *Minerals*, 15(10):1023.

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., Yang, Y., Gupta, S., Majumder, B. P., Hermann, K., Welleck, S., Yazdanbakhsh, A., and Clark, P. (2023). Self-refine: Iterative refinement with self-feedback.

Mansourian, A. and Oucheikh, R. (2024). ChatGeoAI: Enabling geospatial analysis for public through natural language, with Large Language Models. *ISPRS International Journal of Geo-Information*, 13(10):348.

Meta Platforms, I. (2025). Llama 4: Scout & Maverick Models. https://www.llama.com/ or model card at Hugging Face. Released under the Llama 4 Community License, April 5 2025.

OpenAI (2023). ChatGPT (may 13, 2023 version). Large language model.

OpenAI (2025). ChatGPT (august 7, 2025 version). Large language model.

Ovanger, O., Lee, D., Eidsvik, J., Hauge, R., Skauvold, J., and Aune, E. (2025). A statistical study of latent diffusion models for geological facies modeling. *Mathematical Geosciences*, 57(6):1135–1159.

Parquer, M. N., de Kemp, E. A., Brodaric, B., and Hillier, M. J. (2025). Checking the consistency of 3d geological models. *Geoscientific Model Development*, 18(1):71–100.

Pozo, M. F., Saltos, A. G., Llumiquinga, Y. T., Aguirre, K. L., Jara, M. C., and Mejia-Escobar, C. (2025). Geolog-IA: Conversational system for academic theses.

Pyrcz, M. J. and Deutsch, C. V. (2014). *Modeling Principles*, chapter 2, pages 27–151. Oxford University Press, 2 edition.

Ran, X., Xue, L., Sang, X., Pei, Y., and Zhang, Y. (2022). Intelligent generation of cross sections using a conditional generative adversarial network and application to regional 3D geological modeling. *Mathematics*, 10(24):4677.

Ren, A., Wu, L., Xu, J., Xing, Y., Qiu, Q., and Xie, Z. (2025). A deep learning method for 3D geological modeling using ET4DD with offset-attention mechanism. *Computers & Geosciences*, 200:105929.

Song, H. (2024). Geologic interpretation method and system based on vision-language model. US Patent App. 18/610,814.

Thiele, S. T., Jessell, M. W., Lindsay, M., Wellmann, J. F., and Pakyuz-Charrier, E. (2016). The topology of geology 2: Topological uncertainty. *Journal of Structural Geology*, 91:74–87.

https://doi.org/10.5194/egusphere-2025-6545
Preprint. Discussion started: 10 March 2026

Weijermars, R., Waheed, U. b., and Suleymanli, K. (2023). Will ChatGPT and related AI-tools alter the future of the geosciences and petroleum engineering? *First Break*, 41(6):53–61.

Wellmann, F. and Caumon, G. (2018). 3-d structural geological models: Concepts, methods, and uncertainties. *Advances in Geophysics*, 59:1–121.

Woodward, N. B., Boyer, S. E., and Suppe, J. (1989). *Balanced Geological Cross-Sections: An Essential Technique in Geological Research and Exploration*. Short Courses in Geology, Vol. 6. American Geophysical Union.

Wu, G., Wang, H., Zhang, K., Zhu, S., Cui, J., Fan, M., Wang, H., Guo, H., Zhang, G., Zhang, D., Wei, H., and Zhao, S. (2025). GeoProspect: A domain-specific geological large language model with enhanced continual learning. *Neurocomputing*, 650:130801.

Xie, Y., Goyal, A., Wu, X., Yin, X., Xu, X., Kan, M.-Y., Pan, L., and Wang, W. Y. (2024). Coral: Order-agnostic language modeling for efficient iterative refinement.

Xiong, Y., Zuo, R., Luo, Z., and Wang, X. (2021). A physically constrained variational autoencoder for geochemical pattern recognition. *Mathematical Geosciences*, 54(4):783–806.

Xu, Z., Wang, Z., Li, S., Zhang, X., and Lin, P. (2024). Geopredict-llm: Intelligent tunnel advanced geological prediction by reprogramming large language models. *Intelligent Geoengineering*, 1(1):49–57.

Zakrevsky, K. E. (2011). *Geological 3D Modelling*. Earth Resources and Economic Geology. EAGE Publications, Houten, The Netherlands. Product code: EA003, Paperback, 261 pages.

Zhang, H., Xu, J.-J., Cui, H.-W., Li, L., Yang, Y., Tang, C.-S., and Boers, N. (2023). When geoscience meets foundation models: Towards general geoscience artificial intelligence system.

Zhang, H., Xu, J.-J., Cui, H.-W., Li, L., Yang, Y., Tang, C.-S., and Boers, N. (2025). When geoscience meets foundation models: Toward a general geoscience artificial intelligence system. *IEEE Geoscience and Remote Sensing Magazine*, pages 2–41.

Zhang, J., Clairmont, C., Que, X., Li, W., Chen, W., Li, C., and Ma, X. (2024a). Streamlining geoscience data analysis with an LLM-driven workflow. *SSRN Electronic Journal*.

Zhang, Y., Wang, Z., He, Z., Li, J., Mai, G., Lin, J., Wei, C., and Yu, W. (2024b). BB-GeoGPT: A framework for learning a large language model for geographic information science. *Information Processing & Management*, 61(5):103808.

Zhang, Y., Wei, C., He, Z., and Yu, W. (2024c). GeoGPT: An assistant for understanding and processing geospatial tasks. *International Journal of Applied Earth Observation and Geoinformation*, 131:103976.

Zhong, D.-y., Wang, L.-g., Bi, L., and Jia, M.-t. (2019). Implicit modeling of complex orebody with constraints of geological rules. *Transactions of Nonferrous Metals Society of China*, 29(11):2392–2399.