

Review

Not peer-reviewed version

Application of Large Language Models in Geotechnical Engineering: A Movement Towards Safe and Sustainable Future

[Kaustav Chatterjee](#)*, [Mohak Desai](#), [Joshua Li](#)

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0334.v2

Keywords: large language models; transformer architecture; geotechnical engineering; data driven techniques



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

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

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

Review

Application of Large Language Models in Geotechnical Engineering: A Movement Towards Safe and Sustainable Future

Kaustav Chatterjee ^{1,*}, Mohak Desai ² and Joshua Li ³

¹ PhD, School of Civil and Environmental Engineering, Oklahoma State University, Stillwater, Oklahoma, United States of America

² Senior Data Analyst, NIQ, Vadodara

³ Professor and Williams Professorship, School of Civil and Environmental Engineering, Oklahoma State University, Stillwater, OK

* Correspondence: kaustav.chatterjee10@okstate.edu

Abstract

Over the last two decades, there has been a paradigm shift in geotechnical engineering driven by advances in sensing, communication, and data-driven techniques. These advancements enhanced the safety and reliability of geotechnical infrastructure through real-time monitoring and automated decision-making. In recent times, Large Language Models (LLMs) have emerged as advanced data-driven techniques contributing to automated risk assessment of geotechnical infrastructure. LLMs are advanced deep learning models widely used to solve complex numerical problems, analyze large volumes of data, and generate human language. This paper presents a comprehensive review of the application of LLM in geotechnical engineering. The integration of LLMs into geotechnical engineering has demonstrated significant advances in slope stability analysis, bearing capacity computation, numerical analysis, soil-structure interaction, and underground infrastructure. By summarizing the latest research findings and practical applications, this research paper underscores the potential of LLMs to advance and automate various processes in geotechnical engineering. The findings presented in this paper not only provide insights into the current LLM-based geotechnical practices but also emphasize the instrumental role LLM can play in advancing geotechnical engineering, ultimately ensuring a safer and more sustainable future. Lastly, this paper highlights the different LLM capabilities which can be used to empower geotechnical engineers.

Keywords: large language models; transformer architecture; geotechnical engineering; data driven techniques

Introduction

In recent times, the frequency of natural disasters has increased due to various natural and anthropogenic factors. These natural disasters have contributed to the destruction of various types of geotechnical infrastructure, causing slope failures, landslides, soil instability, and foundation failures, often resulting in the loss of property and lives and damage to other infrastructure. The risk of infrastructure failure can be minimized by incorporating real-time monitoring and advanced design techniques.

Conventionally, risk assessment of geotechnical infrastructure is performed manually or using various numerical techniques, such as the finite element method, finite difference method, and discrete element method. However, these manual techniques have shortcomings, including time consumption, high labor costs, and the lack of feasibility for real-time risk assessment. Moreover, numerical techniques have certain drawbacks, such as complexity, limited generalizability, and time

consumption. The problems can be solved by deploying advanced data-driven techniques along with advanced sensing and communication devices for geotechnical risk assessment and monitoring.

Artificial Intelligence (AI) has emerged as a state-of-the-art data-driven technique for simulating human mind and functions, such as problem-solving, learning, reasoning, and perception. By analyzing the dataset, AI understands patterns to make predictions and decisions. With the use of AI, industry tools are transforming from generative tools and healthcare diagnostics to the automatic execution of tasks in manufacturing. Unlike conventional software, AI learns patterns from data to improve its performance.

Machine Learning (ML) is a subfield of AI that focuses on equipping systems with the ability to understand, learn, and improve from experience without being programmed specifically. ML models are trained using databases containing both input and output data for supervised learning and only input data for unsupervised learning. The model learns different patterns from the data, allowing it to make accurate predictions. Common examples of machine learning algorithms are random forest, gradient boosting, decision tree, and support vector machine (Chatterjee et al., 2024, Parajulee et al., 2025). Common applications of machine learning in geotechnical engineering include bearing capacity estimation, slope stability analysis, and soil property prediction. There are certain shortcomings of ML models, including the inability to capture complex non-linear relationships, poor performance on sequential and time-series data, and difficulty processing unstructured data, such as images. These shortcomings can be alleviated using Deep Learning (DL) models.

DL is a specialized subset of ML inspired by the function and structure of the human brain and uses multiple hidden-layer neural networks to mimic its learning process. DL involves the use of an ANN, which comprises interconnected nodes that process and transmit information similar to biological neurons. DL algorithms have proven successful for tasks such as natural language processing, image, and speech recognition. Some popular deep learning architectures include convolutional neural networks, Recurrent Neural Networks (RNNs), and the transformer architecture.

Conventionally, deep learning algorithms such as RNNs and transformer architectures (Vaswani et al., 2017; Ansari et al., 2025; Chatterjee et al., 2026) are used to solve various problems in geotechnical engineering. However, there are certain drawbacks associated with deep learning models, including a lack of understanding of natural languages, poor generalization, and the need for large amounts of training data for model development. The inability of deep learning models to understand human language limits their ability to automate various workflows in geotechnical engineering. Moreover, as deep learning models are developed for specific cases, their generalizability decreases. For instance, a deep learning model trained to predict the factor of safety of one region cannot be used to predict the factor of safety of another region, as it was trained on a limited amount of data. These shortcomings of deep learning models can be addressed by incorporating LLMs into geotechnical engineering.

The novelty of this study lies in summarizing the application of LLMs across different sectors of geotechnical engineering, including slope stability analysis, foundation bearing capacity computation, tunnel and underground infrastructure design, automated numerical modelling, and automated geotechnical site characterization. Moreover, this study outlines the advantages and shortcomings of incorporating LLMs across different geotechnical sectors. LLM automated various processes in geotechnical engineering. There are certain shortcomings of incorporating LLMs in geotechnical engineering, including the need for large volumes of data for model development and the generation of incorrect outcomes due to LLM hallucinations. Lastly, this study contributes to the fields of AI and geotechnical engineering by defining areas for future research and outlining ways LLMs can be integrated to automate tasks performed by geotechnical practitioners. Figure 1 shows a schematic representation of the application of LLM in geotechnical engineering.

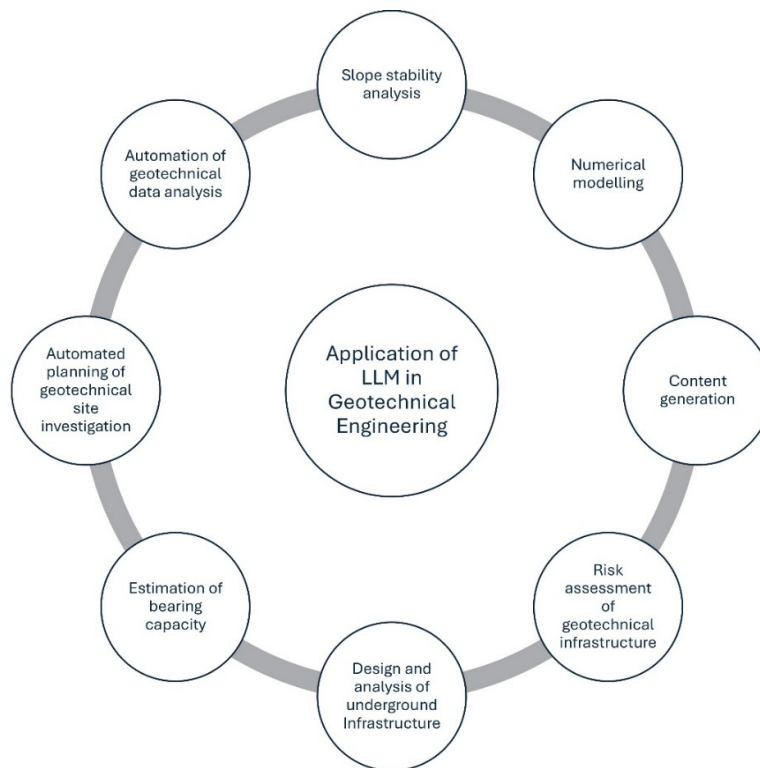


Figure 1. Application of LLM in geotechnical engineering.

This research was initiated by selecting different pairs of keywords, including geotechnical engineering and large language model, geomechanics and large language model, large language model and slope stability, and large language model and bearing capacity analysis. Subsequently, various databases, including ScienceDirect, Google Scholar, the American Society of Civil Engineers database, and Multi Disciplinary Publishing Institution (MDPI) databases, were selected to acquire literature. From various databases, approximately 30 research papers on the application of LLMs across different areas of geotechnical engineering were identified and summarized. Figure 2 shows the adopted methodology used in this research.



Figure 2. Schematic representation of the research.

Large Language Models

Large Language Models (LLMs) have emerged as the next generation of advanced AI models and are increasingly adopted to understand and solve complex problems, analyze large volumes of data across various domains, and generate coherent human language. These models are trained on large volumes of data and, unlike traditional natural language models that were trained on specific domains, these models are designed as task-agnostic architectures with significantly high (hundreds of billions (Kaplan et al. 2020, Workshop et al. 2022, Zhang et al. 2022) number of parameters. Due to their scale and scope, LLMs have also shown emergent capabilities such as autonomous decision-making, reasoning, contextual learning, and planning.

The architectures of most prevalent Large Language Models are based on transformer architecture with encoder-only, decoder-only, or encoder-decoder model variants. Transformers are a type of deep learning architecture developed by Vaswani et al. (2017) for machine translation that can capture long-term dependencies in sequences. Before transformer architectures, the Long Short-Term Memory (LSTM) architecture was the most popular for capturing dependencies in data. However, the LSTM architecture cannot capture long-term dependencies in data and is not suitable for language processing. Another benefit of transformer architecture is faster data processing speed as compared to LSTM architecture due to its ability of parallel processing (Vaswani et al. 2017, Zheng et al. 2020, Islam et al. 2023).

The main components of the transformer architecture are the Encoder and Decoder blocks, along with positional encoding. The encoder block processes the input data to generate a mathematical representation of the input sequence, and the decoder block generates the output sequence from one token at a time. Generally, transformer architectures lack computational units to encode input data positions, but in sequence modelling, positional information is important; positional encoding provides positional information for each data point in the input sequence. Equations 1 and 2 represent the mathematical computation of the positional encoding block.

$$PE(\text{position}, 2e) = \sin(\text{position}/10000^{2e/f}) \quad (1)$$

$$PE(\text{position}, 2e + 1) = \cos(\text{position}/10000^{2e/f}) \quad (2)$$

Where position represents the token of the input data, e represents the dimension of the position index vector, and f represents the dimension of the representation.

The important components of transformer architecture are a multi-head self-attention layer and a position-wise feed-forward network. The multi-head self-attention layer computes relationships between different data points in the input, as shown in Equations 3 and 4. The feed-forward neural network then processes the self-attention layer's output to help the model understand complex data relationships; Equation 5 represents this computation. The dropout layer helps prevent overfitting, residual connections address overfitting, and layer normalization stabilizes network training.

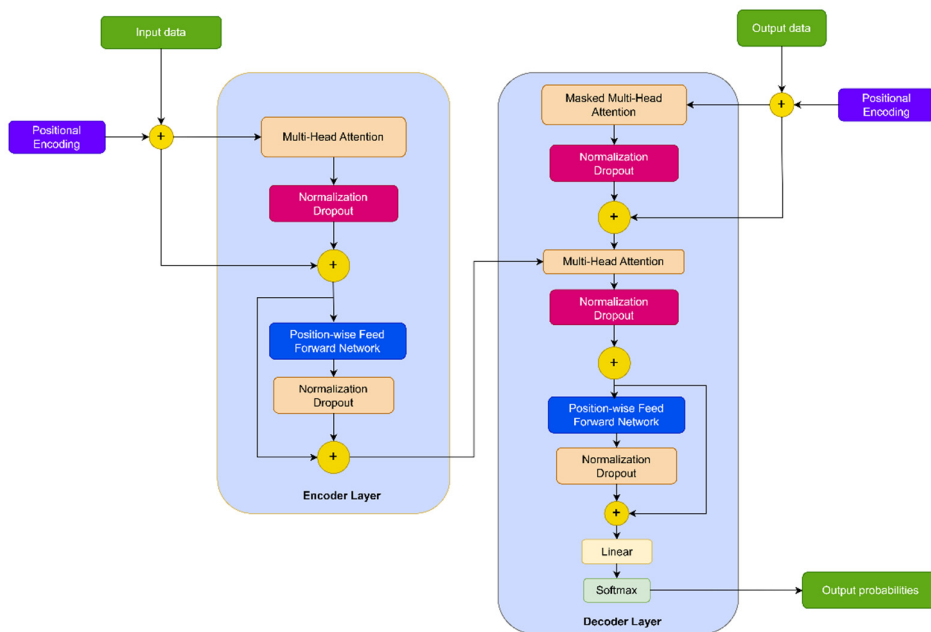
$$\text{Attention}(X) = \sigma\left(\frac{DE^T}{\sqrt{d_s}}\right)F \quad (3)$$

$$D = XW_x; E = XW_y; F = KW_z \quad (4)$$

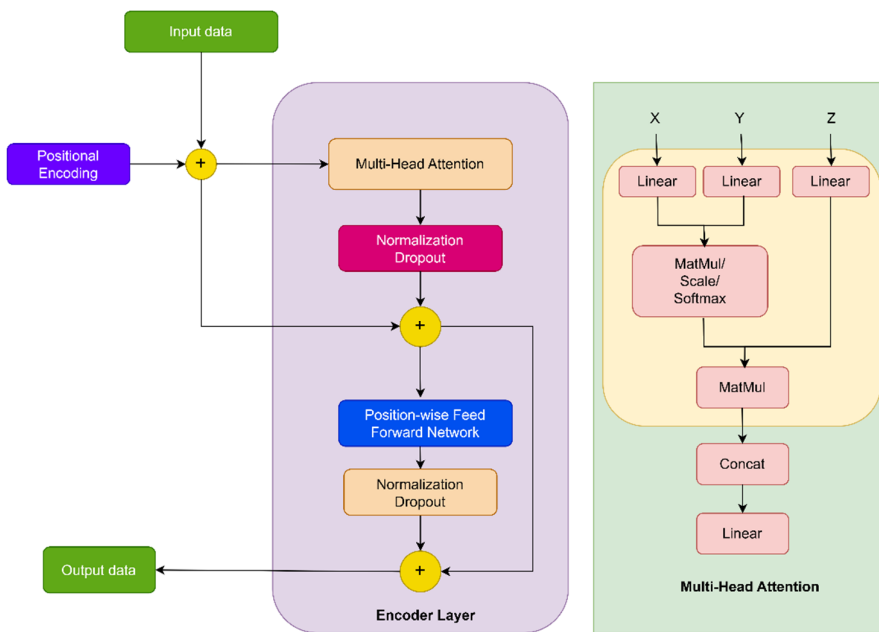
$$FN(x_k) = W_6 \cdot \text{ReLU}(h_i \cdot W_5 + c_6) + c_5 \quad (5)$$

Where X represents the encoded value with positional encoding, σ represents the Softmax function, D , E and F represents the query, key and value vector, respectively. d_s represents the dimension of key vector and W_x , W_y , and W_z represents the weight matrix of query, key and value respectively. x_k represents the input vector at position k , W_5 and W_6 represents the weight matrix of the first and second linear layer respectively, c_5 and c_6 represents the bias vector of the first and second linear layers respectively, and ReLU represents the rectified linear unit activation function.

The LLM model based on transformer architecture can be classified into three categories, namely: (a) encoder model, (b) decoder model, and (c) encoder-decoder model. Figure 3 shows a schematic representation of the three LLM architectures. Encoder-only models (Figure 3b) are generally used for analyzing data and can be used for summarizing borehole logs and geotechnical data reports. Decoder-only models (Figure 3c) are used for prediction and content generation and can generate codes for numerical analysis in geotechnical engineering and for answering questions about different concepts in geotechnical engineering. Encoder-decoder models (Figure 3c) are used for language translation and can be applied to translate geotechnical engineering reports from one language to another.



(a)



(b)

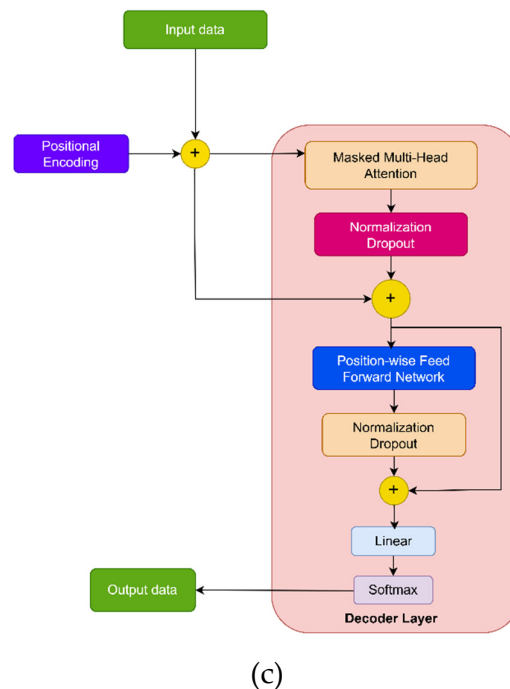


Figure 3. Schematic representation of LLM, (a) Encoder-decoder model, (b) Encoder only model, (c) decoder only model.

Application of LLM in Geotechnical Engineering

Stability Analysis of Slope

Slope stability is an important aspect in geotechnical engineering, critical for stability of highway and railway embankments, bridge abutments, earth dams, mine pits and tailings storage facilities. Conventionally, stability analysis of soil slopes is performed using Swedish Circle / Fellenius Method, Bishop's Simplified Method, Janbu Method and Morgenstern-Price Method. The shortcomings of the conventional techniques are assumption of predefined failure surface, lack of calculation of pore water during analysis, failure to consider the constitutive relationship of soil mass and cannot model deformation during slope failure. These shortcomings were overcome by using finite element analysis for slope stability using different commercial software. The problems associated with the commercial software are the complex procedure of finite element analysis. LLM can facilitate slope stability analysis by providing guidance into that analysis and can also interpret the results of the slope stability analysis. Recent times, researchers adopted LLM for slope stability analysis. Table 1 shows the application of LLM for automating slope stability analysis.

Figure 4 shows a schematic representation of slope stability analysis using LLM. The LLM receives a prompt from users, processes it, generates code, and uses that code to produce results. The user reviews the results and provides feedback to the LLM. This process continues until the result appears correct to the user.

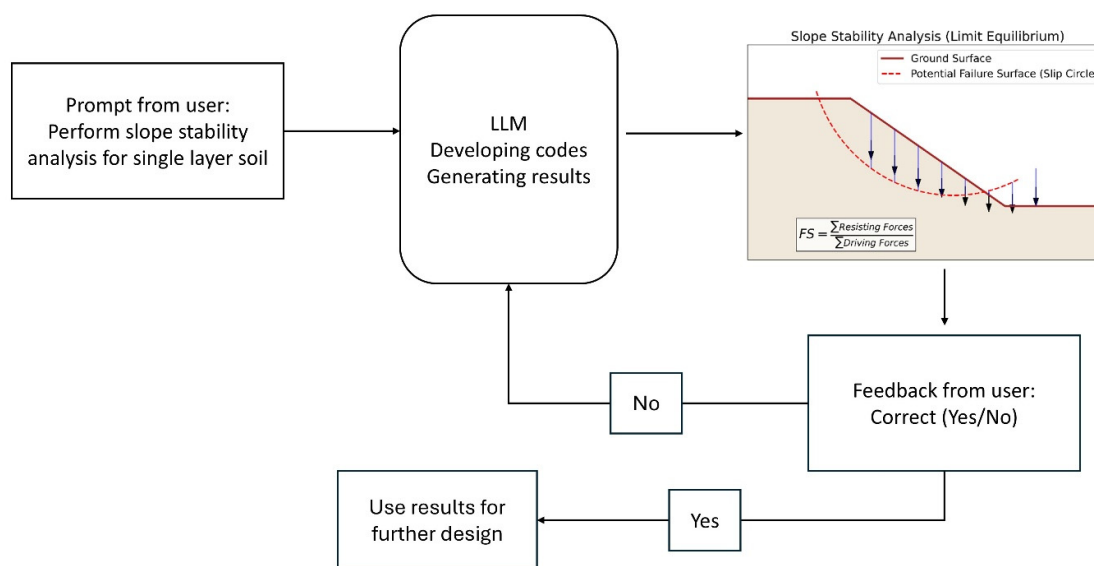


Figure 4. Schematic representation of slope stability analysis using LLM.

Table 1. Application of LLM in slope stability.

Authors	Scientific contribution	LLMs
Kwak and Won (2025)	LLM for python code computation for seepage analysis and slope stability and prompt driven framework for coupling seepage results with slope stability calculation	ChatGPT ChatGPT-o1
Kim et al. (2024)	Generation of MATLAB functions to calculate the Factor of Safety (FS) using ChatGPT. Results validated against GeoStudio SLOPE/W	ChatGPT-4.0, BERT, T5
Wu et al. (2024)	Use of LLM to analyze slope photographs and generate descriptive text of slope features relevant to stability and to predict collapse risk as a percentage	ChatGPT

Kim et al. (2024) used ChatGPT-generated MATLAB code to identify critical failure surfaces and calculate the Factor of Safety (FS) using the Fellenius method of slices. Additionally, ChatGPT was prompted to generate MATLAB code for solving seepage flow using the Finite Difference Method. The code computed the hydraulic head distribution and produced flow nets comparable to those produced by commercial software (GeoStudio SEEP/W). The results were validated against GeoStudio SLOPE/W, showing accurate identification of critical failure surfaces and achieved FS values of 1.630. The major advantage of this study is that ChatGPT was able to logically establish programming sequences, including the definition of variables and domains, the formulation of governing equations, iterative operations, and convergence checks, and the visualization of results. In addition, ChatGPT's outputs were consistent with results from commercial software such as GeoStudio SEEP/W and SLOPE/W.

Xu et al. (2025) introduced a GPT-4o-based multi-GeoLLM, a multimodal, multi-agent MML framework that integrates text and image inputs to automate geotechnical tasks such as footing

design, bearing capacity, and settlement analysis, and to generate GPT-assisted MATLAB code for slope stability evaluation. In addition, the model generates design drawings using Python-based logic and equations derived from standard codes. The proposed model, Multi-GeoLLM, achieved perfect accuracy of 1.0 across 60 multimodal cases (text, image, and text-image sub-tests) in footing design cases, whereas it achieved 0.97 accuracy with 100 textual cases.

Wu et al. (2024) conducted a study that used photo analysis and textual reasoning to automate visual inspection of slopes and assess landslide risk. In addition, ChatGPT was employed for site-similarity prediction, simulation-parameter recommendation, and site grouping based on seismic hazard. The primary role of the LLM in these applications was to group sites with similar seismic characteristics, auto-generate Python code for spatial analysis and plotting, extract guidance from the LIQCA manual using Retriever-Augmented Generation, recommend parameter values for given soil types, and compare scatter plots of clay properties across sites. In this study, it was observed that GPT successfully grouped sites based on HVSr curves and spatial data. In addition, when location data was added, GPT-generated Python code improved clustering accuracy to match expert recommendations. Also, in site-similarity prediction, GPT's rankings of similar sites were generally consistent with those of a hierarchical Bayesian model.

Kwak and Won (2025) attempted to integrate an LLM, specifically ChatGPT, into advanced geotechnical analyses by developing a framework for seepage-induced slope stability assessment. Their study showed that ChatGPT can generate Python code for seepage modeling, slope stability calculations using Bishop's simplified method, and the coupling of both analyses, achieving factors of safety within 1.86% margin of the industry standards (SEEP/W and SLOPE/W). It was observed that the LLM incorporated optimization techniques, automated phreatic line extraction, and reduced computational time by up to 70% through adaptive algorithms, demonstrating the potential of LLMs to make geotechnical workflows more efficient, accessible, and sustainable. This work underscores how LLMs can make complex numerical modeling accessible, reduce reliance on expensive software, and accelerate decision-making. This is a significant step toward embedding AI-driven automation in geotechnical engineering for sustainable infrastructure design. Additionally, the study highlights a human-in-the-loop approach to refining prompts when ChatGPT misinterprets tasks (e.g., correcting the slice angle calculation). This demonstrates that while LLMs can automate engineering workflows, expert oversight remains critical for accuracy and robustness.

Design and Analysis of Tunnels and Underground Infrastructure

Tunnels are underground structures that pass through soil and rock and are important components of the transportation infrastructure, facilitating the movement of people and freight by highways and railways. Traditional methods of tunnel stability analysis and underground engineering are performed using analytical and empirical techniques developed by researchers, as well as numerical methods such as finite element and finite difference methods. The problems with empirical methods include limited generalizability, as these methods may not be applicable across different geologic conditions, and the neglect of the stress-strain behavior of rock or soil. Unlike empirical methods, numerical techniques can be applied across different geologic conditions and account for the stress-strain behavior of geomaterials. However, numerical techniques are complex and time-consuming. LLMs can overcome the limitations of empirical and analytical methods, as they are trained on large volumes of data. Researchers used an LLM to predict various parameters in underground technology.

Table 2. Application of LLM for designing tunnels and underground infrastructure.

Authors	Scientific contribution	LLMs
Wu et al. (2026)	Tunnel rock mass integrity prediction by integrating multi-modal data (images, radar, drilling, text) using a generative LLM	GPT-4
Wu et al. (2025a)	Tunnel face stability evaluation by integrating LLM with multimodal knowledge graph (MMKG)	GPT-4o, DeepSeek-R1, Ali-Qwen, Doubao, Keling, Yuanbao, Gemini 1.5
Njock et al. (2025)	Tunnel structural failure risk assessment into levels (Low, Medium, High, Critical) using natural-language inputs in transformer-based LLM called DistilBERT	GPT-4
Hu et al. (2025)	Integration of LLM into Tunnel Boring Machine (TBM) operations for human-machine collaboration, intention recognition, and decision transparency	Qwen1.5-32B
Mehrishal et al. (2025)	Demonstration of practical integration of LLMs into tunnel geotechnical workflows, automated tunnel face mapping and rock mass characterization	GPT-4
Xu et al. (2024b)	Integration of LLM into tunnel advanced geological prediction by reprogramming LLMs	BERT, GPT-2, LLaMA

Xu et al. (2024) presented a study on application of LLMs in geotechnical engineering, especially for tunnel advanced geological prediction by creating GeoPredict-LLM framework. This study also put in place a new approach to reuse pretrained LLMs like BERT, GPT-2, LLaMA for tunnel geological estimation by reprogramming them instead of fine-tuning. For this, multimodal geotechnical data are converted into language-compatible data so that LLMs can work on numerical information by employing its pretrained reasoning capability. Geological, geophysical, and drilling data are first merged via knowledge graph embedding and after that transformed into linguistically structure which can be processed by LLMs. By transforming geological prediction to language-based task, the approach improves accuracy (BERT, GPT-2, LLaMA > 90%) and reduces computational cost which enables improved decision making for underground and tunnel engineering.

Wu et al. (2025) demonstrated how LLMs are emerging as an important tool for tunnel construction. This study showcased a tunnel-specific LLM (Tunnel-GPT, Tunnel-DeepSeek, Tunnel-AliQwen, etc.) driven multimodal framework that integrates images, videos, drilling data, GPR signals, and geological sketches into a unified knowledge graph to automate tunnel-face stability prediction. Additionally, LLMs were also employed to create high-fidelity synthetic Rock-Mass images to improve dataset balance and to increase the diversity of geological conditions. By combining LLMs for synthetic rock-mass image generation with computer-vision models and a structured knowledge graph, the framework achieves high accuracy (up to 96%) under complex geological conditions and reduces reliance on manual inspections. Overall, these innovations make LLM-driven multimodal systems an important technology for achieving a more sustainable, real-time evaluation of tunnel-face stability.

In the same year, Tiwari et al. (2025) demonstrated a semantic AI framework, GeoSemantica, that uses fine-tuned LLMs to assess seismic soil liquefaction risk. The key application of LLMs is to binary-classify soil liquefaction occurrence under seismic loading. The LLM examines the semantic history derived from geotechnical and seismic inputs to determine whether liquefaction is possible at the site. GeoSemantica translates geotechnical parameters, such as effective stress, soil type, SPT-N value, and seismic loading, into domain-informed natural language to summarize geotechnical reasoning. This allows LLM to record interactions between soil properties and seismic demand. The GeoSemantica LLM achieved 75.0% accuracy, 81.5% F1, and a remarkably high recall, outperforming other LLMs. This study shows that LLM approaches can give more reliable decision-making in geotechnical earthquake engineering.

Another research, Hu et al. (2025), presented the application of LLM by developing an LLM-based intelligent assistant for autonomous Tunnel Boring Machine (TBM) tunneling. This research combined an LLM with domain-specific knowledge and a multi-agent framework to enable human-machine collaboration in complex underground construction scenarios. Moreover, by combining a stepwise LLM with RAG, the framework can predict operator intention, support decision-making, and monitor anomalies during tunneling operations. Case studies of metro tunnel projects demonstrate that LLM-based assistants notably enhance system transparency, reduce manual intervention, and improve operational safety. From a sustainability perspective, this work demonstrates how LLMs can enable more efficient, reliable geotechnical construction by optimizing automated operations and minimizing manual errors.

Mehrishal et al. (2025) presented an AI-driven framework, Tunnel Rapid AI Classification (TRaiC), that demonstrates the role of LLMs in geotechnical engineering workflows, particularly in underground engineering. This research combined computer vision-based discontinuity detection, 360° tunnel face imaging, 3D digital twin generation, and the RAG-LLM system to automate interpretation and standardized reporting. In this framework, the LLM acts as an intelligent geotechnical assistant, blending multimodal inputs such as images, discontinuous data, and historical tunnel data to provide rock mass descriptions and Rock Mass Rating (RMR) values aligned with engineering standards. By minimizing reliance on manual tunnel-face mapping, this LLM-based system improves efficiency and safety while reducing human involvement in risky environments and situations.

The most recent study on the application of LLMs in tunnel engineering was conducted by Wu et al. (2026), who developed a Tunnel Rock Integrity Prediction GPT (Tunnel RIP-GPT) for tunnel rock mass integrity assessment. The study showed that a GPT-4-based LLM can effectively combine diverse multimodal data, including tunnel face images, geological sketches, ground-penetrating radar outputs, drilling parameters, and physico-mechanical properties within a semantic framework. Traditional ML models like CNN and transformer-based models struggle with multimodal integration, but LLM applies attention-based language-driven interaction to achieve end-to-end prediction of rock mass integrity, and the accuracy numbers achieved are more than 90%. Moreover, the study includes diffusion-based image generation to address data imbalance and enables prompt-based interactions for tunnel engineers, reducing dependence on site testing.

LMM-Assisted Bearing Capacity Calculation

Bearing capacity is an important concept in geotechnical engineering and is used for the design of shallow foundations and deep foundations. Conventionally, bearing capacities are estimated using standard penetration test values, soil types from borehole logs, and empirical equations developed by researchers. There are certain shortcomings of empirical equations, including the assumed failure mechanism of soil, not considering the stress-strain relationship of soil, inadequate representation of soil stratification, and ignoring the stress history of soil. These shortcomings can be solved by performing a finite element analysis of foundations. However, the shortcomings of finite element analysis include complex analysis and time-consuming calculations. These shortcomings can be overcome by using an LLM to design foundations. LLM models developed for bearing capacity

estimation are trained on large volumes of data and may exhibit better generalizability across different conditions. Table 3 summarizes the application of LLM for bearing capacity calculation. Figure 5 shows a schematic representation of bearing capacity estimation for shallow foundations using an LLM. Users provide a prompt to the LLM, the LLM processes it, and generates code to implement the design. The user reviews the results, and once satisfied, the design can be implemented.

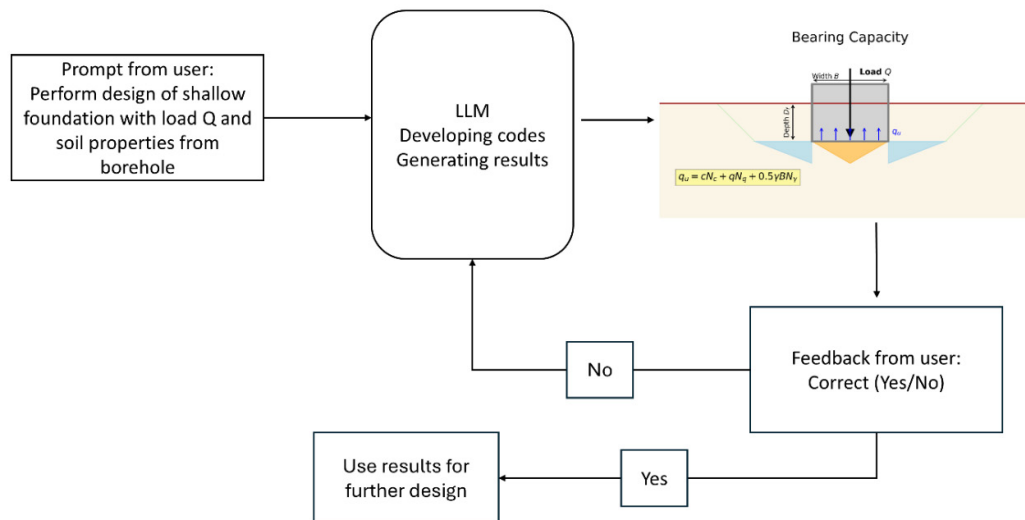


Figure 5. Schematic representation of design of shallow foundation using LLM.

Table 3. Application of LLM for bearing capacity estimation.

Authors	Scientific contribution	LLMs
Kim et al. (2025b)	Automated Python code generation for calculating pile vertical bearing capacity according to API RP 2A	ChatGPT, GPT-4o, GPT-o1
Xu et al. (2024a)	Introduction of hybrid prompt engineering approach for geotechnical tasks such as bearing capacity calculation and settlement estimation by using a domain-specific framework, GeoLLM	Gemini-pro, GPT-4, GLM-4, Qwen family
Kumar (2024)	Engineering calculation workflows formation via LLM + tools (ReAct framework) for computing bearing capacity	GPT-3.5, GPT-3.5-turbo

Xu et al. (2024) developed a Gemini-pro-based GeoLLM model to estimate bearing capacity and settlement for a single pile. Main tasks involve extracting design parameters from geotechnical texts and performing calculations in accordance with European, Chinese, and American design codes. In addition, this study evaluates various LLMs, including Gemini-pro, GPT-4, GLM-4, and the Qwen family, for accuracy in extracting geotechnical parameters and reliability in performing engineering calculations. The study demonstrates that LLMs with >100B parameters are suitable for high-precision engineering tasks. The main advantage of this model is its remarkable text comprehension and human-like responses, enabled by its transformer architecture. Also, the GeoLLM model attained high precision (up to 0.988) for intelligent geotechnical designs. The following year, Kim et al. (2025)

presented a study demonstrating the use of ChatGPT to automate the calculation of vertical pile bearing capacity in accordance with API RP 2A design standards. The key application of LLM in this study is to generate Python code for calculating pile vertical bearing capacity, to read and understand API RP 2A design standards, and to extract equations, parameter limits, and tabulated coefficients. The study highlighted that ChatGPT successfully generated valid computational workflows for shaft friction, end bearing capacity, and penetration depth estimation through prompt interaction. LLM-assisted code generation remarkably excels direct numerical computation by LLMs and minimizes arithmetic errors. This approach has been proven to deliver consistent geotechnical design workflows by reducing repetitive manual calculations, thereby promoting sustainable geotechnical problem-solving.

Virtual Assistance, Knowledge support, Content Generation and Problem Solving

In earlier times, the major sources of knowledge for geotechnical engineers were books, journal papers, lecture notes, and videos. It was cumbersome and time-consuming for engineers to learn various geotechnical engineering concepts. These problems can be solved by using an LLM to retrieve information on various geotechnical engineering concepts. LLM chatbots are developed based on different scientific literature and can answer basic to advanced-level questions in geotechnical engineering. Geotechnical engineers can leverage LLM for quick reference to questions. Table 4 summarizes the application of LLM for providing knowledge support in geotechnical engineering. Figure 6 shows the process of numerical modelling code generation using LLM. The input from user is processed using decoder-based architecture LLM, subsequently the code is generated.



Figure 6. Content generation using LLM.

Table 4. Application of LLM for providing content generation in geotechnical engineering.

Authors	Scientific contribution	LLMs
Soranzo (2025)	Educational content generation and automated grading system development using LLM	ChatGPT-4.0
Kim et al. (2025b)	Training LLM with interactive workflow via prompt engineering to interpret and implement design standard	ChatGPT, GPT-4o, GPT-o1
Tophel et al. (2025)	Demonstration of general-purpose LLMs as an AI tutor for geotechnical engineering when augmented with RAG via APIs	GPT-4, LLaMA-3
Babu et al. (2025)	Assessment of LLM as virtual assistant for fundamental, practical, and advanced technical topics	ChatGPT, Copilot, Gemini
Reddy & Janga (2025)	Evaluation of LLM capabilities and limitations for real geotechnical tasks such as literature review,	GPT-3.5 GPT-4 Microsoft Bing

	report drafting, coding, data analysis, and conceptual understanding	Google Bard Meta LLaMA
Xu et al. (2025)	Retrieval of Building Information Modelling data using language instructions with the use of “BIMS-GPT” framework	GPT-4o
Chen et al. (2024)	Evaluation of the capabilities of GPT-4 in geotechnical education, problem-solving assistance and interactive tutoring	GPT-4
Kumar (2024)	Conceptual reframing of LLM as language-based reasoning engines and formalization of prompt engineering as a methodological requirement	GPT-3.5 GPT-3.5-turbo
Zhang et al. (2025)	Use of LLM with carefully curated prompt as a decision-making tool with agent-based architecture	ChatGPT-4o

Chen et al. (2024) performed a study to address a major research gap by systematically evaluating GPT-4's capabilities in geotechnical education and problem-solving. The study includes a question bank of 391 questions covering soil mechanics, permeability, shear strength, slope stability, and bearing capacity. In this study, GPT-4 is envisioned as an AI tutor that can provide personalized instruction to students, correct errors in responses, explain reasoning steps, and serve as a feedback mechanism. Also, GPT-4 was applied to solve textbook-based geotechnical problems, including calculations for stresses, void ratios, and bearing capacities. GPT-4 achieved 28.9% accuracy with baseline performance without guidance, 34% accuracy when reasoning steps are requested, and 67% accuracy when domain-specific instructions are provided.

Liu and Shi (2025) conducted a study demonstrating the capability of LLM (GPT-4) to automatically extract critical information, such as geological conditions, laboratory test results, and engineering recommendations, from conventional geotechnical reports. Moreover, GPT-4 can parse general project metadata, subsurface and hydrogeologic conditions, design recommendations, spatial artifacts such as site maps and boring logs, and laboratory tests, and stream these outputs into AR-based 3D visualizations for on-site decision support. This practice reduces the time and expertise taken for manual data processing, reduces human errors, and promotes data-driven decision-making and positions LLMs – especially GPT-4 as a key enabler of sustainable geotechnical practices by supporting safer field operations and improving the overall lifecycle management of infrastructure projects.

Soranzo (2025) demonstrated in a study that LLMs like ChatGPT-4.0, DistilBERT, and MiniLM, when fine-tuned on geotechnical textbooks and domain-specific texts, can generate high-quality educational content, automate grading of technical responses and reports, and support consistent decision-making aligned with established soil mechanics and geotechnical design principles. In this study, GPT 4.0, BERT, and MiniLM were employed for generating geotechnical question-answers, creating synthetic student answers, computing cosine similarity for grading, and classifying student answers in Grades 1 to 5. LLM-based grading systems, supplemented by cosine similarity and retrieval-augmented generation, have improved the evaluation of open-ended geotechnical questions, achieving up to 98% accuracy after fine-tuning and surpassing traditional similarity-based methods. Moreover, a web-based, threshold-powered tool for embedding and grading was developed that instantly evaluates student responses and provides feedback. In sum, LLMs deliver near-human consistency with ~97.5–98.3% accuracy on fine-tuned open-ended grading and ~71.4% on full technical reports, while offering scalable, low-effort deployment and immediate feedback loops.

In the same year, Babu et al. (2025) conducted a study that included ChatGPT, Microsoft Copilot, and Google Gemini across various geotechnical concepts, such as slope stability, frost action, and cross-anisotropy, and rated their performance as fair, good, and poor. The primary contribution of this study is a domain-specific evaluation of general-purpose LLMs as virtual assistants for fundamental, practical, and advanced technical topics. The study showed that LLMs can assist engineers with conceptual understanding, preliminary analysis, and literature review by providing fluent explanations of soil mechanics problems. While LLMs have strong potential to assist with geotechnical tasks, some limitations, such as misattributed references, incorrect technical generalizations, and failure to contextualize site-specific geotechnical conditions, have been observed. Overall, LLMs are trustworthy decision-support tools that can be employed to increase efficiency, reduce repetitive efforts, and support sustainable development.

Recent studies have demonstrated that LLMs are intelligent knowledge support systems. For example, Tophel et al. (2025) demonstrated the application of GPT-4 and LLaMA-3 as AI educators for undergraduate geotechnical engineering, emphasizing the RAG framework. By merging geotechnical literature with formula repositories via an API, this research demonstrated that LLMs can improve accuracy and reliability in solving geotechnical topics such as consolidation, shear strength, and stress analysis. A GPT-4-based LLM achieved nearly 95% accuracy, showcasing the success of blending LLMs with geotechnical knowledge from the literature. Furthermore, this study underscores the use of LLMs as a supplementary resource similar to textbooks or solution manuals. Through these applications, this study demonstrates that domain-adapted LLMs can serve as scalable, 24/7 knowledge-support tools.

Reddy and Janga (2025) explored AI adoption through a global survey of geotechnical and geoenvironmental professionals, demonstrating that LLMs are primarily used for literature review, technical content preparation, code generation, and data interpretation. Moreover, LLMs have the potential to support sustainable geotechnical engineering practices by enabling efficient analysis of large geotechnical reports and reducing the time required for manual tasks, such as report preparation and data visualization. Apart from these advantages, LLMs also have disadvantages, such as hallucinations, numerical inaccuracies, and a lack of engineering judgment, which, at this point, make LLMs unsuitable for final design decisions. This research concludes that, rather than in decision-making, LLMs can act as intelligent assistants to support decision-making, ultimately reducing human effort and enabling data-driven decision-making in geotechnical engineering.

Risk Assessment of Geotechnical Infrastructure

Risk assessment in geotechnical engineering is performed using different stochastic techniques, such as Monte Carlo simulation, to determine the probability of failure of geotechnical infrastructures. In the present era, researchers are using LLMs to assess the risk of different infrastructure systems. Table 5 shows the application of LLM for risk assessment of geotechnical infrastructure. Figure 7 shows a schematic representation of the risk assessment process using an LLM. The user input to the LLM includes site investigation parameters, sensor data, and instructions for performing the analysis. Based on the available data, the LLM performs Monte Carlo simulations and finite element analyses to generate the probability of failure. The probability of failure can be used in decision-making and hazard mitigation.

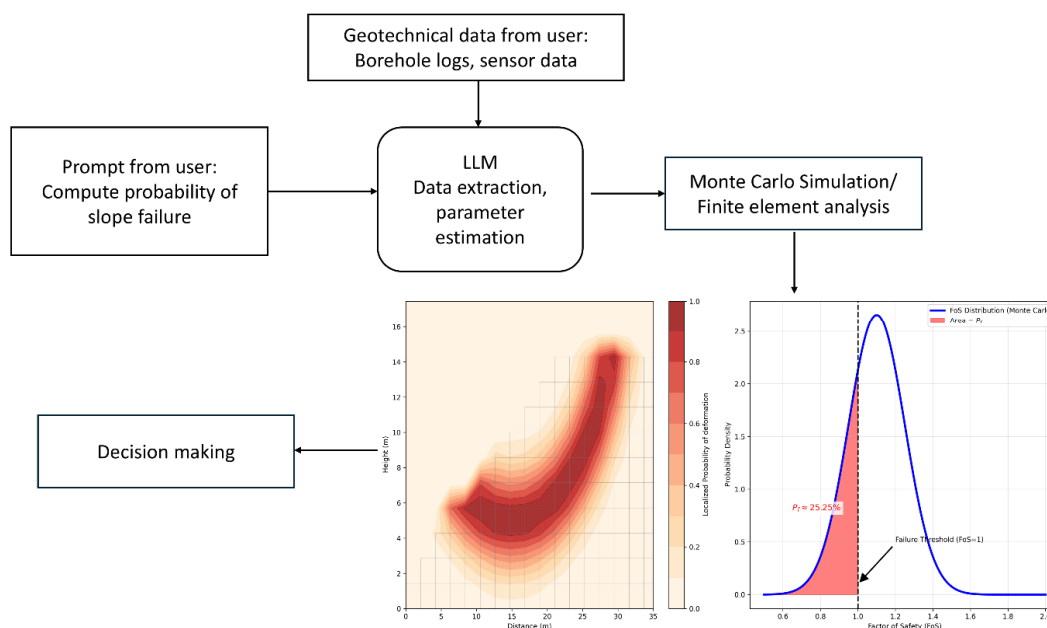


Figure 7. Schematic representation of risk assessment of geotechnical infrastructure.

Table 5. Risk assessment of geotechnical infrastructure.

Authors	Scientific contribution	LLMs
Kamran et al. (2025)	Integration of LLM and prompt engineering in geotechnical risk/rockburst prediction by generating accurate and context-specific outputs	Google Gemini
Pang et al. (2025)	Application of LLM-based agentic AI for post-landslide geotechnical investigations	GPT-3.5-Turbo, GPT-4o
Areerob et al. (2025)	Demonstration of LLM to integrate visual cues extracted from landslide imagery and to perform multi-step geotechnical reasoning	GPT-4, GPT-3.5, LLaMA-2-13B, Alpaca-13B, LaVIN, QLoRA
Njock et al. (2025)	Tunnel structural failure risk assessment using LLM by allowing engineers to query tunnel risk using natural language	GPT-4
Tiwari et al. (2025)	LLM based framework for seismic soil liquefaction risk assessment and structured geotechnical and seismic data conversion into natural language	GPT-4, Gemini Pro, Claude 2.1, Amazon Titan

Njock et al. (2025) presented a study on how LLM can be operationalized for geotechnical risk assessment. The authors develop DistilBERT-TunnelRisk to enable natural language-driven prediction of structural failure risk in shield tunnels. By converting conventional geotechnical inputs such as geological conditions and groundwater levels into question-answer pairs, the model enables engineers to query tunnel risk through conversational text rather than through structured numerical interfaces. The model achieves high predictive accuracy (precision/recall/F1 up to 0.96–1.0) and outperforms general-purpose LLMs like GPT-4 and DeepSeek in domain-specific reasoning. Overall,

this research represents an advancement in applying LLMs to geotechnical engineering tasks, including excavation stability, slope failure assessment, and foundation risk evaluation.

In the same year, Areerob et al. (2025) integrated an LLM with multimodal AI in their study on geotechnical hazard interpretation, particularly expert-level landslide image analysis. The study linked aerial imagery with LLM-based reasoning to recreate the tacit decision-making processes conventionally done by experienced geotechnical engineers. By advancing both a VQA-LLM hybrid framework and an end-to-end multimodal LLM (MLLM), the authors proved how LLMs can be utilized for causal interpretation and future risk assessment of slope failures from visual data. Additionally, the major focus of this study is on the digitalization of expert geotechnical knowledge captured via verbal commentary and structured using LLMs. The outcome demonstrates that LLM-driven systems can provide geologically relevant interpretations and risk insights comparable to those of human experts, highlighting the strong potential of LLMs as decision-support tools in geotechnical engineering. This option is fast, easy, and scalable for landslide assessment.

Another study in 2025 by Pang et al. (2025) examined the reconstruction of landslides and the automation of post-landslide investigation using LLM-based agentic AI. In this research, an LLM was combined with RAG to extract engineering-relevant information, and a multimodal LLM was integrated with fine-tuned vision models, such as YOLO, to estimate landslide geometry from site images. By using pre-trained foundation models and CoT prompting, the suggested framework reduces reliance on large databases and heavy manual effort, two major drawbacks in traditional geotechnical analysis. Results from LLMs applied to historical landslide cases in Hong Kong show that summaries and geometric estimates are consistent with professional forensic reports. This highlights the potential of LLM-based agentic AI to achieve greater efficiency and scalability in hazard investigation, supporting quick risk assessment, better decision-making, and improved planning.

LLM-Driven Automation of Numerical Modelling

Numerical modelling of geotechnical infrastructures is performed to determine the stability of slopes, bearing capacity estimation, settlement of infrastructures, and design of tunnels and underground infrastructures. Numerical modelling of infrastructure is performed using finite difference, finite element, and discrete element methods. Although these finite element and finite difference models are very accurate, these techniques have shortcomings, including complexity, time consumption, and the need for manual interpretation to analyze results. LLMs, on the other hand, can perform fast analysis and interpret results without human intervention. Researchers used an LLM to automate the numerical modelling of infrastructure and the interpretation of results. Table 6 shows the application of LLM for numerical modelling in geotechnical engineering. Figure 8 shows a schematic representation of numerical modelling for slope stability analysis using an LLM. The LLM receives the prompt from the user and generates code to solve numerical problems. There is a feedback loop between the LLM agent and the numerical solver, which enables the LLM to improve the code and debug errors.

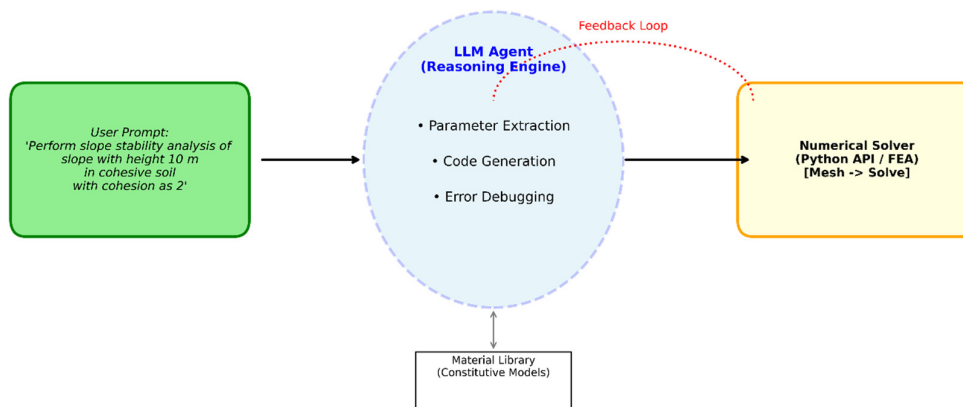


Figure 8. Schematic representation of numerical modelling using LLM.**Table 6.** Application of LLM for automating numerical modeling in geotechnical engineering.

Authors	Scientific contribution	LLMs
Kim et al. (2025a)	Use of ChatGPT for Finite Element Analysis and hydro-mechanically coupled problems	ChatGPT o1
Bekele (2025)	Numerical simulation management by natural language using GeoSim.AI	No specific algorithm

Bekele (2025) introduced GeoSim.AI, which demonstrates how LLMs can reshape computational geomechanics through numerical simulations, enabling them to be managed via natural language. GeoSim.AI uses LLMs as its central processing unit to translate natural-language or image inputs into full geomechanical simulation scripts for tools such as ADONIS, HYRCAN, PLAXIS, and FLAC. Moreover, this study showcases slope stability modeling in ADONIS and HYRCAN using text-only prompts and combined image-and-text prompts. GeoSim.AI automates repetitive setup tasks, allowing researchers to focus more on geomechanical behavior rather than software operations. Overall, GeoSim.AI's ability to translate natural language and visual inputs into fully structured numerical models makes it efficient for geotechnical design.

Kim et al. (2025) conducted a study on the use of ChatGPT for Finite Element Analysis of soil-structure interaction and coupled hydro-mechanical problems. This work demonstrates how LLMs can autonomously generate executable FE code for Single-field problems, such as 1D consolidation using Terzaghi's equation, and Mixed-field problems, such as coupled displacement-pore pressure formulations. Additionally, this work addresses three benchmark problems: 1D consolidation (fluid mass diffusion), Differential settlement of a strip footing, and Gravity-driven seepage in unsaturated soil. By validating GPT-generated FE codes against analytical solutions and experimental data, this study provides a proof-of-concept for integrating AI into computational geomechanics workflows. From this study, it was observed that while using advanced libraries like FEniCS, ChatGPT required minimal code revisions and passed verification tests quickly, which is its primary advantage. Whereas a low-level programming environment like MATLAB failed even after multiple prompt augmentations, requiring direct human intervention.

Kamran et al. (2025) demonstrate an integration of LLMs and Generative AI for geotechnical risk prediction, specifically focusing on rockburst hazards in underground construction. By leveraging Google Gemini's multimodal (text, code, audio, images, PDFs, and video) reasoning and prompt-engineering-driven automation, the authors show how LLMs can independently generate, refine, and validate Python code for complex geotechnical analyses, which can transfer traditionally manual and time-intensive processes into adaptive, data-driven workflows. Furthermore, the LLM was used to generate Pie charts for rockburst intensity distribution, pairwise scatter plots for variable relationships, and 3D plots for factor analysis and clustering results. The research highlights how LLMs help engineers in shifting from reactive safety measures to predictive, sustainable risk-mitigation strategies by enabling automated data processing, factor analysis, clustering, and ML-based intensity forecasting with high accuracy. This work represents an emerging direction in which LLMs act not only as conversational assistants but also as intelligent analytical partners capable of enhancing underground risk assessment.

Automation in Geotechnical Site Investigation Planning

Conventionally, geotechnical site investigation planning is performed manually based on project requirements. Manual methods are time-consuming and costly. Researchers are developing LLM-based techniques to automate geotechnical site investigation planning. Qian and Shi. (2025) presented an LLM(GPT-4O) empowered study that advances geotechnical engineering workflows by integrating RAG and agentic human-machine collaboration. Their work demonstrates how LLMs

can automate key components of site investigation, including information retrieval from multi-source site investigation design codes, automated borehole layout planning, and rapid geological characterization from multimodal data. The prepared LLM model can automatically generate borehole spacing, depth, and layout schemes in accordance with regional codes, providing near-real-time, code-compliant site investigation plans. By proposing a Multihop-RAG framework that is capable of accurately extracting domain-specific clauses and generating feasible sampling schemes, the study showcases the potential of LLMs to enhance efficiency, reduce human error, and support real-time, risk-informed decision-making. This contribution highlights an important movement toward sustainable geotechnical engineering by enabling digital transformation, improving resource optimization during site investigations, and fostering interpretable AI-assisted geotechnical analysis. Table 7 shows the application of LLM for automating geotechnical site investigation planning. Figure 9 shows the automated site investigation planning process using an LLM. The LLM receives parameters from the user and project information, and generates borehole plans that account for various constraints and requirements. The design engineer reviews the output, and there is a feedback loop between the user and the LLM to optimize the design.

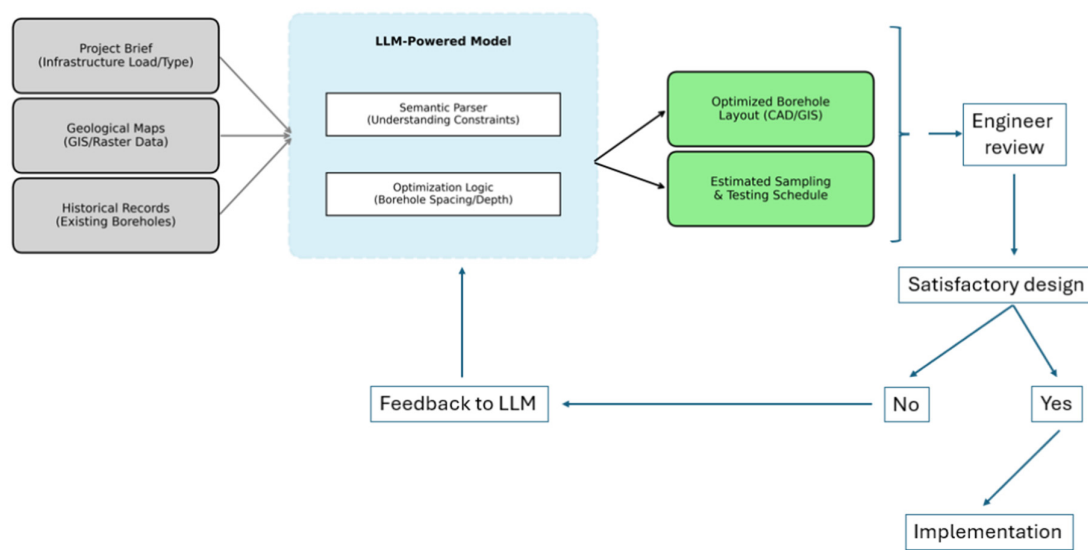


Figure 9. Schematic representation of automated site investigation planning using LLM.

Table 7. Application of LLM for geotechnical site investigation planning.

Authors	Scientific contribution	LLMs
Fan et al. (2026)	Domain specific adaptation of LLMs in geotechnical engineering tasks such as automated borehole layout, lithology classification and site characterization	GPT-3.5, GPT-4, GPT-4o, GPT-o1
Xu et al. (2025)	Automation in design process for geotechnical tasks by using Multi-GeoLLM- a multimodal, multi-agent LLM framework	GPT-4o
Qian and Shi (2025)	Automated site planning, interpretation of geotechnical literature and design codes comparison with LLM	GPT-4o

Liu and Shi (2025)	Integration of LLM data in AR based 3D visualization for on-site decision support	GPT-4
Wu et al. (2025b)	Development of agentic AI for computation for geotechnical tasks with natural language as a formal interface	GPT-3, GPT-4, PaLM, Gemini, LLaMA
Li and Shi (2025)	Automatic creation of geological cross-sections from sparse borehole data using LLM	ChatGPT-4.0

Li and Shi (2025) presented a study on the automatic generation of geological cross-sections from sparse borehole data using ChatGPT-4.0. This research presented that LLMs can understand geotechnical reasoning from a few-shot textual examples without model retraining. The model developed a prompting strategy by integrating Few-shot examples to teach domain rules, Chain-of-Thought (CoT) reasoning to enforce multistep logic, and Self-consistency sampling. Two tunneling and reclamation case studies were validated in this study, in which LLMs were employed to determine stratigraphic boundaries and generate 2D geological cross-sections. This framework achieved an accuracy of ~77%, demonstrating that LLMs can reduce reliance on expert manual interpretation and improve consistency and efficiency. This proves LLMs as a powerful mechanism for scalable subsurface modelling.

In the same year, Wu et al. (2025) conducted a study on the use of LLM-based agentic AI in geotechnical engineering, demonstrating its ability to transform labor-intensive, expert-advised workflows. This LLM agent was used for geotechnical site planning, landslide investigation and post-event analysis, liquefaction analysis, and shield tunnel safety evaluation. Additionally, LLMs were used to extract design clauses from multilingual geotechnical codes and guidelines using RAG, and to automatically generate geological cross-sections from sparse borehole data using agentic workflows. Moreover, this study introduced a natural language-based geotechnical computation that uses natural language as a formal interface for site characterization, design reasoning, risk evaluation, and regulatory compliance checks. These automations ultimately decrease repetitive manual efforts and create proactive risk management, which ultimately creates sustainable geotechnical engineering practices.

LLM-Driven Workflow Automation in Geotechnical Data Analysis:

Recent studies have demonstrated that LLMs are domain-specific workflow controllers rather than text generators. Zhang et al. (2025) conducted LLM-driven research in which LLMs serve as action agents to understand natural-language queries, retrieve geoscientific details, and execute analytical and visualization tasks via external tools. The LLM natural language queries are transformed into machine-readable API parameters for the OpenMindat API. By merging LLM as a connecting layer between users and domain APIs, the workflow reduces programming queries and improves consistency of data-driven analyses. Moreover, this study demonstrates that domain-specific fine-tuning of LLMs is not essential for complicated geoscience workflows. Instead, well-designed prompts and tool schemas will enable generic-purpose LLMs to work productively in specialized engineering contexts.

Challenges in LLM-Driven Geotechnical Engineering

Although LLMs have found large-scale applications across different sectors of geotechnical engineering, there are certain challenges in applying them to geotechnical design and analysis. Some challenges in implementing LLMs in geotechnical engineering include the need for large volumes of data, hallucinations in LLMs, and the requirement for extensive computational resources.

LLM generally provides better generalizability of results than conventional statistical models, machine learning models, and deep learning models. Better generalizability comes at the cost of

requiring a large volume of data. Generating huge volumes of data is difficult in geotechnical engineering. For instance, LLM models developed for slope stability or embankment construction analysis are based on limited sets of subsurface data due to the high cost of acquiring them. The limited number of subsurface affects the generalizability of the models. One potential solution to this issue is to generate data using various data augmentation techniques.

LLM sometimes hallucinates and provides incorrect solutions and codes for solving geotechnical engineering problems. Incorrect answers from LLMs may mislead geotechnical practitioners, leading to incorrect estimates of bearing capacity and the factor of safety for slopes and embankments. One solution to this problem is to cross-check the LLM-based solution against the literature.

Another shortcoming of incorporating an LLM-based approach in geotechnical engineering is prompt engineering. Prompt engineering is the process of structuring and designing instructions for LLMs. Adequate prompt engineering reduces hallucination, reduces computational times, and improves automation. Kumar (2024) presented a study demonstrating the stochastic parrot problem in LLMs, where LLMs can produce fluent but incorrect or misleading outputs. Moreover, this study provided a clear experimental example in which GPT produced incorrect soil classification results when asked for direct answers. However, when authors applied chain-of-thought prompting, the model correctly followed all logic, including the liquid limit threshold, plasticity index computation, and A-line comparison, and correctly generated a classification. Furthermore, this study is among the first to formalize prompt engineering as a methodological requirement rather than a user-convenience tool.

Finally, one major problem associated with the development of LLM-based geotechnical tools is the requirement for huge computational resources as they are developed on huge volumes of data. The requirement for substantial computational resources increases costs and limits their use to edge devices.

Summary

This paper underscores the growing significance of LLMs in geotechnical engineering across various applications, including bearing capacity and slope stability calculations, tunnel structural assessment, LLMs as virtual assistants, knowledge support, failure risk assessment, automated numerical modeling, and automated site investigation. LLMs are increasingly used to generate automated MATLAB/Python code for seepage flow analysis, failure surface detection, and slope stability evaluation. The research underscores the human-in-the-loop approach for clarifying prompts when ChatGPT misinterprets tasks (e.g., correcting slice angle calculation). LLM results, especially those from ChatGPT, were consistent with those from commercial software such as GeoStudio SEEP/W and SLOPE/W. Furthermore, advanced models such as Multi-GeoLLM achieved a perfect accuracy of 1.0 with 60 multimodal cases and enable automated design drawings.

In addition to slope stability, LLMs are transforming tunnel engineering by converting complex geological and drilling data into a structured format, enabling high-precision estimates of tunnel-face and rock-mass stability. Furthermore, models like Tunnel-GPT and GeoPredict-LLM utilize multimodal inputs such as images, GPR signals, drilling logs, and geological sketches to automate and streamline forecasting with high accuracy.

For bearing capacity estimation, several limitations of traditional methods can be overcome by using LLMs for foundation design, as LLMs (such as GeoLLM) are trained on large datasets and exhibit better generalizability across conditions. Recent studies even claim the automation of pile-bearing capacity calculations using LLMs, generating a trustworthy workflow that can minimize manual tasks.

Geotechnical engineering is shifting from textbook-based literature to LLM-based virtual assistance, which can provide instant support. Research in 2024 and 2025 highlights the capability of GPT-4 and similar models to solve soil mechanics problems, extract insights from geotechnical reports, grade student responses, and even support 3D AR-based visualization. With fine-tuning and RAG frameworks, LLMs achieved a high precision of 95–98%, but limitations such as hallucinations

and a lack of engineering judgment persist. Overall, LLMs can serve as digital tutors and knowledge partners, always available.

Future Roadmaps

The future research related to LLM application in geotechnical engineering can be categorized into three different components: (a) development of LLM with improved architecture, achieving better accuracy and performance, (b) development of LLM for better generalizability, and (c) developing LLM-based approaches that did not exist before. An LLM with a better architecture can be developed for complex tasks that require nonlinear stress-strain behavior or large-strain behavior, including predicting particle flow during landslides and slope failures, and the consolidation of compressible soils. Moreover, advanced LLMs should be developed to solve multiphysics problems in geotechnical engineering, including the coupled thermal, hydraulic, and mechanical behavior of energy piles and the hydraulic and mechanical behavior of geomaterials during rainfall-induced landslides.

One of the major problems in geotechnical engineering is the acquisition of large volumes of data, which hinders the generalizability of LLMs. The generalizability of LLMs in geotechnical engineering can be improved by generating data with different data augmentation techniques or by using generative AI. Another approach to increasing generalizability is to train an LLM on data from regions around the world. For instance, an LLM trained to analyze slope stability in one region can be trained on data from another region to improve generalization.

LLM can be integrated with modern technologies, such as digital twins, to enhance the safety and reliability of geotechnical infrastructure. A digital twin is a concept consisting of sensing devices, virtual models, and a communication system. The sensing device acquires real-time information from the geotechnical infrastructure; the virtual model functions as a replica of the real infrastructure; and the communication system enables real-time information transfer between the model and the infrastructure. LLM can be integrated into the digital twin infrastructure, serving as a decision-making tool to improve the safety and reliability of infrastructure.

Integration of LLM for Geotechnical Practitioners

LLM can be employed to advance and automate the activities that practicing engineers perform. Figure 10 lists some of the tasks that can be automated using LLM. These tasks include scanning geotechnical reports and borehole logs, answering geotechnical engineering questions, and reviewing geotechnical instrumentation data and preliminary design and analysis. Conventionally, scanning information from geotechnical reports and borehole logs requires a considerable amount of time; applying an LLM for this purpose can save time for geotechnical engineers. Geotechnical engineers can use LLMs to obtain quick answers on various geotechnical engineering concepts; however, those answers should be verified by experts for accuracy before using them for decision-making in a project. Engineers can use an LLM to review sensor data from geotechnical infrastructure to perform a real-time risk assessment. Moreover, engineers can use LLM to perform preliminary design of geotechnical infrastructure. Overall integration of LLMs for performing geotechnical engineering tasks will reduce the time and expense required for these tasks and reduce errors.

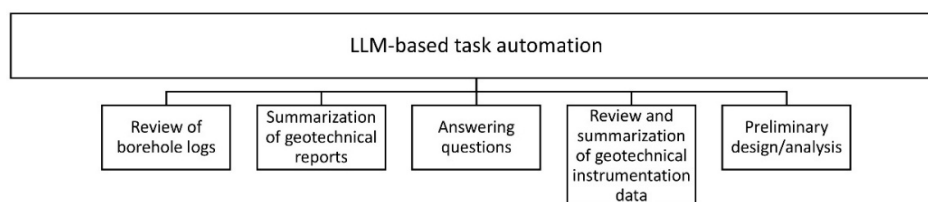


Figure 10. LLM-based task automation.

Reference

- Ansari, F., Chatterjee, K., Li, J. Q., Wang, K., & Golalipour, A. (2025). Multi-Object Pavement Surface Feature Detection with CNN and Transformer Deep Learning Architecture. In *Airfield and Highway Pavements 2025* (pp. 350-359).
- Areerob, K., Nguyen, V. Q., Li, X., Inadomi, S., Shimada, T., Kanasaki, H., ... & Okatani, T. (2025). Multimodal artificial intelligence approaches using large language models for expert-level landslide image analysis. *Computer-Aided Civil and Infrastructure Engineering*.
- Bekele, Y. W. (2025). GeoSim. AI: AI assistants for numerical simulations in geomechanics. *arXiv preprint arXiv:2501.14186*.
- Chatterjee, K., Vivanco, D., Yang, X., & Li, J. Q. (2024). Enhancing pavement performance through balanced mix design: A comprehensive field study in Oklahoma. In *International Conference on Transportation and Development 2024* (pp. 511-522).
- Chatterjee, K., Li, J. Q., Ansari, F., Munna, M. R., Parajulee, K., & Schwennesen, J. (2026). Hybrid LSTM-Transformer Models for Profiling Highway–Railway Grade Crossings. *Journal of Transportation Engineering, Part A: Systems*, 152(2), 04025138.
- Chen, L., Tophel, A., Hettiyadura, U., & Kodikara, J. (2024). An investigation into the utility of large language models in geotechnical education and problem solving. *Geotechnics*, 4(2), 470-498.
- Fan, L., Liu, F., & Chen, C. (2026). Domain adaptation of large language models for geotechnical applications. *Solid Earth Sciences*, 11(1), 100285.
- Hu, M., Gao, H., Mi, Q., Wu, B., Lu, J., & Liu, Y. (2025). Bridging the Information Gap in Smart Construction: An LLM-Based Assistant for Autonomous TBM Tunneling. *Smart Cities*, 8(6), 212.
- Islam, S., H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, and W. Pedrycz. 2023. "A comprehensive survey on applications of Transformers for deep learning tasks." *Expert Systems with Applications*, 122666.
- Kamran, M., Faizan, M., Wang, S., Han, B., & Wang, W. Y. (2025). Generative AI and Prompt Engineering: Transforming Rockburst Prediction in Underground Construction. *Buildings*, 15(8), 1281.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.
- Kim, D., Kim, T., Kim, Y., Byun, Y. H., & Yun, T. S. (2024). A ChatGPT-MATLAB framework for numerical modeling in geotechnical engineering applications. *Computers and geotechnics*, 169, 106237.
- Kim, S., Kim, D., & Youn, H. (2025). Automating vertical bearing capacity calculations using python: Prompt engineering of ChatGPT on API RP 2A. *Developments in the Built Environment*, 21, 100628.
- Kim, T., Yun, T. S., & Suh, H. S. (2025). Can ChatGPT implement finite element models for geotechnical engineering applications?. *International Journal for Numerical and Analytical Methods in Geomechanics*, 49(6), 1747-1766.
- Kumar, K. (2024). Geotechnical parrot tales (gpt): Harnessing large language models in geotechnical engineering. *Journal of Geotechnical and Geoenvironmental Engineering*, 150(1), 02523001.
- Kwak, J., & Won, J. (2025). Application of ChatGPT in seepage-induced slope stability. *KSCE Journal of Civil Engineering*, 100457.
- Li, H., & Shi, C. (2025). Few-shot learning of geological cross-sections from sparse data using large language model. *Geodata and AI*, 2, 100010.
- Liu, Z., & Shi, Y. (2024, August). Leveraging Large Language Models and Augmented Reality to Enhance the Understanding of Geotechnical Report. In *International Conference on Information Technology in Geo-Engineering* (pp. 116-124). Cham: Springer Nature Switzerland.
- Mehrishal, S., Leem, J., Kim, J., Shao, Y., Kang, I. S., & Song, J. J. (2025). Tunnel Rapid AI Classification (TRaiC): An Open-Source Code for 360° Tunnel Face Mapping, Discontinuity Analysis, and RAG-LLM-Powered Geo-Engineering Reporting. *Remote Sensing*, 17(16), 2891.
- Njock, P. G. A., Yin, Z. Y., Xu, H. R., & Zhang, N. (2025). Structural failure risk assessment of shield tunnel using large language model. *Tunnelling and Underground Space Technology*, 165, 106882.
- Pang, H., Lo, M. K., Leung, Y. F., & Wu, S. (2025). Reconstruction of landslide events using LLM-based Agentic AI with multimodal data. Available at SSRN 5340542.

- Parajulee, K., Chatterjee, K., & Li, J. (2025). Leveraging original equipment manufacturer vehicle sensor data for enhanced roadway safety. *International Journal of Pavement Research and Technology*, 1-18.
- Reddy, K. R., & Janga, J. K. (2023, October). Utilization of Generative Artificial Intelligence in Geotechnical and Geoenvironmental Engineering. In *International Conference on Environmental Geotechnology, Recycled Waste Materials and Sustainable Engineering* (pp. 241-267). Singapore: Springer Nature Singapore.
- Suresh Babu, A., Fyaz Sadiq, M., Aydin, C., Velasquez, R., & Izevbekhai, B. (2025). Evaluation of Large Language Models as Geotechnical Virtual Assistant. In *Geotechnical Frontiers 2025* (pp. 48-58).
- Qian, Z., & Shi, C. (2025). Large language model-empowered paradigm for automated geotechnical site planning and geological characterization. *Automation in Construction*, 173, 106103.
- Soranzo, E. (2025). Large language models for automated grading in geotechnics. *Machine Learning and Data Science in Geotechnics*, 1(1), 124-144.
- Tiwari, A., Gupta, A. K., & Rawat, S. (2025). Soil, Seismic Risk, and Semantics: Ai-Driven Tool for Resilient Infrastructure Planning. *Authorea Preprints*.
- Tophel, A., Chen, L., Hettiyadura, U., & Kodikara, J. (2025). Towards an AI tutor for undergraduate geotechnical engineering: a comparative study of evaluating the efficiency of large language model application programming interfaces. *Discover Computing*, 28(1), 76.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Workshop, B., Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., ... & Bari, M. S. (2022). Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.
- Wu, C., Huang, H., & Ni, Y. Q. (2025). Evaluation of tunnel rock mass integrity using multi-modal data and generative large model: Tunnel rip-gpt. Available at SSRN 5348429.
- Wu, C., Huang, H., Yu, Z., Zhang, Y., Chen, S., & Chen, J. Automatic evaluation of tunnel face stability based on generative large language models and multimodal knowledge graphs. Available at SSRN 5870723.
- Wu, S., Otake, Y., Mizutani, D., Liu, C., Asano, K., Sato, N., ... & Yoshikawa, R. (2025). Future-proofing geotechnics workflows: accelerating problem-solving with large language models. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 19(2), 307-324.
- Wu, S., Shi, C., Leung, Y. F., Otake, Y., Konishi, C., Zhou, M., ... & Nakamura, T. (2025). Perspectives: LLM agents reshaping the foundation of geotechnical problem-solving. *Geodata and AI*, 4, 100036.
- Xu, H. R., Zhang, N., Yin, Z. Y., & Njock, P. G. A. (2025). GeoLLM: A specialized large language model framework for intelligent geotechnical design. *Computers and Geotechnics*, 177, 106849.
- Xu, H. R., Zhang, N., Yin, Z. Y., & Njock, P. G. A. (2025). Multimodal framework integrating multiple large language model agents for intelligent geotechnical design. *Automation in Construction*, 176, 106257.
- Xu, Z., Wang, Z., Li, S., Zhang, X., & Lin, P. (2024). GeoPredict-LLM: Intelligent tunnel advanced geological prediction by reprogramming large language models. *Intelligent Geoengineering*, 1(1), 49-57.
- Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., ... & Zettlemoyer, L. (2022). Opt: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*.
- Zhang, J., Clairmont, C., Que, X., Li, W., Chen, W., Li, C., & Ma, X. (2025). Streamlining geoscience data analysis with an LLM-driven workflow. *Applied Computing and Geosciences*, 25, 100218.
- Zheng, Y., X. Li, F. Xie, and L. Lu. 2020, May. "Improving end-to-end speech synthesis with local recurrent neural network enhanced Transformer." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. pp. 6734-6738. Barcelona, Spain.

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