
GEOLIQUEFY: A SEMANTIC AI PLATFORM FOR GEOTECHNICAL RISK ASSESSMENT

Achyut Tiwari*
Geoliquefy
Bengaluru, India
achyut@geoliquefy.com
&
achyuttiwari@icloud.com

Ashok Kumar Gupta
Geoliquefy
Bengaluru, India
&
Department of Civil Engineering
Jaypee University of Information Technology
Solan, HP 173234
ashok.gupta@juit.ac.in

Saurabh Rawat
Geoliquefy
Bengaluru, India
&
Department of Civil Engineering
National Institute of Technology, Delhi
Plot No. FA7, Zone-P1, GT Karnal Road, Narela, Delhi-110036, INDIA
saurabh.rawat@nitdelhi.ac.in

June 22, 2025

ABSTRACT

Soil liquefaction during seismic events poses significant risks to infrastructure and human safety. Traditional machine learning (ML) models often struggle with the sparse, nonlinear, and heterogeneous data typical in geotechnical engineering, as these characteristics hinder their ability to learn generalizable patterns and capture the complex interactions among soil parameters. This paper introduces a novel approach that leverages large language models (LLMs) to interpret geotechnical data by transforming structured datasets into semantic narratives. By doing so, we improve the generalizability, interpretability, and reliability of liquefaction risk assessments. A working MVP is available for public testing at <https://app.geoliquefy.com>.

1 Introduction

Geotechnical engineering relies heavily on empirical data from soil investigations, which are often site-specific, sparse, and difficult to generalize [1, 2]. Conventional models—like support vector machines or decision trees—are constrained by their inability to extract contextual dependencies between features such as soil composition, groundwater depth, and seismic loading [3, 4, 5, 6, 7]. Key challenges include:

- **Data Sparsity:** Site investigations rarely yield more than a few hundred labeled instances. This limits the performance of data-hungry models.
- **Nonlinear Interactions:** Soil behavior under seismic loads is governed by complex, nonlinear interactions not easily captured in tabular form.
- **Expert Rule Dependence:** Many traditional methods still rely on empirical charts or domain rules, which are hard to encode into ML systems.

Recent advances in LLMs offer an opportunity to represent these multifaceted relationships in a more expressive, contextualized format [8, 9, 10].

2 Language Models as Interpretive Frameworks

Transforming geotechnical datasets into structured textual narratives allows us to capitalize on the interpretive capabilities of LLMs. For example, a sample entry like:

"At 10 meters depth, the soil is silty sand with an SPT-N value of 12, saturation of 90 %, and effective overburden stress of 150 kPa under a PGA of 0.25g."

This transformation allows the LLM to:

- **Model Cross-Feature Interactions:** Transformer-based models can attend to distant and nonlinear dependencies within the data.
- **Enable Few-Shot Learning:** LLMs pretrained on broad corpora adapt effectively to small geotechnical datasets.
- **Provide Explainable Outputs:** Generated rationales and confidence scores align with the needs of risk-sensitive domains.

LLMs such as GPT, LLaMA, and Mistral, fine-tuned on domain-specific corpora, offer a flexible foundation for data-to-insight pipelines in geotechnical engineering, as demonstrated by performance metrics like accuracy, sensitivity to edge cases, and alignment with expert rationales.

3 Encoding Geotechnical Semantics

A core contribution of GeoLiquefy is a domain-specific serialization pipeline that transforms structured geotechnical data into natural language inputs. Our method includes:

- **Template-Based Encoding:** Uniform text generation for soil layers, including material type, density, SPT values, saturation, and confining stress.
- **Domain-Aware Augmentation:** Enriching inputs with regional seismicity, CPT/VS profiles, and groundwater tables.
- **Fine-Tuned LLMs:** Model enhanced over a particular chosen data set.

This textual encoding bridges the gap between numerical representation and geotechnical semantics, enabling better generalization across geologies—for instance, by allowing models trained on one region’s data to perform well in different seismic or soil contexts.

4 Integration with Geotechnical Pipelines

GeoLiquefy’s architecture supports seamless integration into existing geotechnical workflows:

- **Input Layer:** Accepts CSV or JSON formats with soil profile data.
- **Transformer Inference Layer:** Processes serialized narratives through fine-tuned LLMs.

5 Web Deployment and Public MVP

A public MVP of the GeoLiquefy platform is accessible at:

<https://app.geoliquefy.com>

The platform allows users to:

- Upload site data for instant liquefaction risk analysis.
- Visualize risk levels by depth and location. (upcoming)
- View LLM-generated justifications for each prediction. (upcoming)
- Download formatted risk reports for documentation. (upcoming)

6 Conclusion

GeoLiquefy reimagines geotechnical risk assessment through the lens of language modeling. By embedding soil profiles into semantically rich textual formats, LLMs provide context-aware, generalizable predictions that surpass the limitations of traditional ML methods. This approach brings AI and domain science into synergy, unlocking new possibilities in infrastructure safety, planning, and disaster mitigation.

References

- [1] Ronald D Andrus and Kenneth H Stokoe. Liquefaction resistance of soils from shear-wave velocity. *Journal of Geotechnical and Geoenvironmental Engineering*, 126(11):1015–1025, 2000.
- [2] Robert Kayen, Robb ES Moss, Eric M Thompson, Raymond B Seed, Kemal Onder Cetin, Armen Der Kiureghian, Yasuo Tanaka, and Kohji Tokimatsu. Shear-wave velocity-based probabilistic and deterministic assessment of seismic soil liquefaction potential. *Journal of Geotechnical and Geoenvironmental Engineering*, 139(3):407–419, 2013.
- [3] Anthony TC Goh. Seismic liquefaction potential assessed by neural networks. *Journal of Geotechnical Engineering*, 120(9):1467–1480, 1994.
- [4] Anthony TC Goh and SH Goh. Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data. *Computers and Geotechnics*, 34(6):410–421, 2007.
- [5] Ankit Kumar, Rajesh Kumar Singh, and Jitendra Singh Yadav. A smarter approach to liquefaction risk: harnessing dynamic cone penetration test data and machine learning for safer infrastructure. *Frontiers in Built Environment*, 10:1495472, 2024.
- [6] Ahmad Mahmoudi and Alireza Derakhshani. Liquefaction potential assessment of soils using machine learning techniques: A review. *International Journal of Geomechanics*, 23(7):04023134, 2023.
- [7] Ching-Yinn Lee and Shuh-Gi Chern. Application of a support vector machine for liquefaction assessment. *Journal of Marine Science and Technology*, 21(3):318–324, 2013.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [9] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
- [10] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Nee-lakantan, Pranav Shyam, Sasank Girish, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.