Journal of Geo-Energy and Environment
http://dx.doi.org/10.62762/JGEE.2026.158449

PERSPECTIVE

From Data-Driven to Geology-Informed Learning: Geometry-Aware Graph Transformer Learning for Geo-Energy Systems

Jinghan Li[1,*]

[1] Lincoln University College, Petaling Jaya 47301, Selangor, Malaysia

Abstract

Geo-energy systems involve heterogeneous subsurface properties, irregular spatial structures, and complex interactions. Conventional data-driven models either treat observations independently or rely on graph connectivity without explicitly encoding relative geometry, which limits their ability to distinguish spatially different configurations. This study develops a Geometry-Aware Graph Transformer (GAGT) that integrates node attributes, graph structure, and relative geometric information through a learnable geometric bias in the attention mechanism. A controlled synthetic experiment with 300 irregularly distributed three-dimensional nodes compares a multilayer perceptron (MLP), a Graph Transformer (GT), and GAGT under identical settings. Across 10 random seeds, GAGT achieved a mean absolute error (MAE) of 0.1569, a root mean square error (RMSE) of 0.2152, and a coefficient of determination (R^2) of 0.8678, reducing MAE and RMSE by 35.77% and 33.95%, respectively, relative to GT. In this controlled setting, explicit geometry therefore carries predictive information that graph connectivity alone does not provide. GAGT can serve as a geometry-aware representation layer between purely data-driven learning and more comprehensive geology-informed modeling; validation with real geological data and physically constrained simulations is still required.

Keywords: geo-energy systems, geometry-aware learning, graph Transformer, subsurface modeling, geology-informed artificial intelligence.

1 Introduction

Geo-energy systems, including geothermal reservoirs, subsurface energy storage, geological carbon storage, and petroleum-bearing formations, are characterized by strong spatial heterogeneity and complex geological structures. Their responses depend not only on local properties such as permeability, porosity, pressure, and temperature, but also on how subsurface components are arranged relative to one another. A useful computational model must therefore represent both local attributes and the spatial organization through which these attributes interact.

Submitted: 25 August 2026
Revised: 19 September 2026
Accepted: 29 September 2026
Published: 22 October 2026

Vol. 3, **No.** 1, 2027.
10.62762/JGEE.2026.158449

***Corresponding author:**
✉ Jinghan Li
jinghan.phdscholar@lincoln.edu.my

Citation
Li, J. (2026). From Data-Driven to Geology-Informed Learning: Geometry-Aware Graph Transformer Learning for Geo-Energy Systems. *Journal of Geo-Energy and Environment*, 3(1), 72–83.

© 2026 by the Author. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Conventional geo-energy analysis relies on numerical simulators built on conservation laws, constitutive relationships, and geological models. These simulators give physically interpretable solutions, but they become expensive when runs must be repeated for forecasting, uncertainty analysis, optimization, or operational decision-making. Deep learning and machine learning offer an alternative way to approximate the mapping between system inputs and responses [2, 3]. Deep learning has been used for porous-media flow [4], geothermal resource development [5, 6], decision support for geological CO_2 storage [7], and a range of reservoir and subsurface prediction tasks [8–10].

ML applications in geo-energy have also moved toward more specialized geological and petrophysical tasks, including reservoir-quality prediction [8, 9], shear-wave-velocity estimation [10, 11, 20], fracture-density and fracture-network characterization [12–15], and graph-based production forecasting, subsurface flow optimization, and CO_2 plume migration prediction [16–19]. Further work has applied graph attention networks to production indicator prediction in fractured-vuggy reservoirs [21]. Data-driven machine learning has also been applied to wettability assessment for hydrogen geo-storage [22]. Most of these studies formulate prediction as a mapping from measured or engineered features to target variables, so the spatial organization of the geological system is rarely represented explicitly.

This becomes a problem when the geological configuration changes. A model may learn statistical relationships that hold for a particular permeability distribution, boundary condition, fracture arrangement, or operating regime, yet the spatial relationships behind the response may differ between geological realizations. The same governing equations can produce different responses on different spatial configurations, because permeability distribution, geological boundaries, inter-location distances, and connectivity all affect pressure propagation, flow pathways, and energy transfer. The question is therefore not only predictive accuracy but whether the learned representation retains the spatial information that matters when the configuration changes.

Graphs are a natural representation for such spatially distributed systems. Nodes can stand for grid cells, wells, monitoring locations, or geological subdomains, and edges encode relationships among them. Graph neural networks (GNNs) propagate information along these edges so that each node representation incorporates its neighborhood [23, 24].

Connectivity alone, however, does not specify geometry. Two graphs can share the same connectivity pattern while differing in physical distance, coordinate displacement, orientation, or spatial scale; an edge says that two nodes interact, not how they are positioned relative to each other. In the subsurface, where spatial configuration influences the propagation of pressure, heat, and fluids, this distinction matters. Geometric deep learning has emphasized structured representations for non-Euclidean data [23], and graph-based geo-energy work such as pore-scale flow prediction has shown the value of representing spatial relationships explicitly [25]. These observations motivate a representation that keeps graph connectivity while also exposing relative geometry to the model.

Graph Transformers are a convenient starting point. The Transformer introduced attention-based learning of relationships among input elements [26], and later work adapted attention to graph representation learning [27, 28]. In a conventional Graph Transformer the attention score is computed from node representations within the graph structure. It can learn different interaction strengths among neighbors, but quantities such as inter-node distance and coordinate displacement do not enter the score explicitly.

The gap addressed here is how relative geological geometry can be built into Graph Transformer attention for geo-energy systems in a transparent and computationally light way, without replacing graph connectivity.

We propose a Geometry-Aware Graph Transformer (GAGT) that adds relative-geometric descriptors, namely inter-node distance and coordinate displacement, to the attention score. The model thus combines three kinds of information: local node attributes, graph connectivity, and relative spatial geometry. The formulation is kept deliberately simple so that the effect of geometry can be separated from that of graph structure and node features.

Here, *geology-informed* refers only to the explicit use of geological and spatial descriptors in the learned representation. The model does not solve governing equations or enforce physical constraints, which distinguishes it from physics-informed learning,

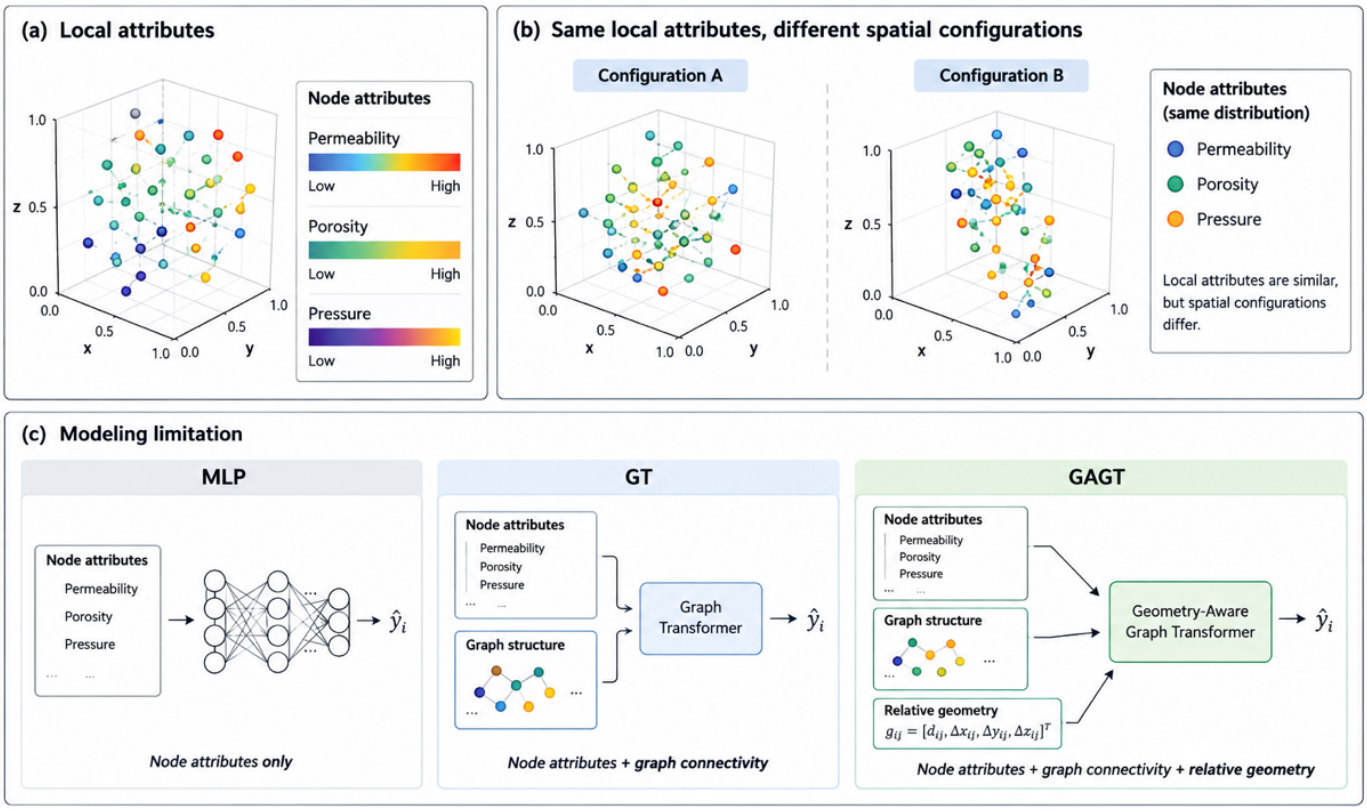

Figure 1. Problem definition and motivation for geometry-aware learning. (a) Local node attributes in an irregular three-dimensional subsurface setting. (b) Spatial configurations with similar local attributes but different relative spatial relationships. (c) Information inputs used by the MLP, Graph Transformer (GT), and Geometry-Aware Graph Transformer (GAGT).

where physical laws or differential-equation residuals enter the training objective [29, 30]. GAGT is intended as a structural component that can later be combined with physical constraints or domain-informed graph construction.

GAGT is evaluated in a controlled synthetic experiment against an MLP and a conventional GT, using MAE, RMSE, and R^2. The comparison is designed to isolate the incremental contribution of graph structure and of explicit geometry; it does not reproduce a field-scale reservoir, and the proximity-based graph does not encode hydraulic connectivity, fracture topology, or permeability-weighted interactions.

The study addresses three research questions:

(1) How can geological geometric information be incorporated into a Graph Transformer for geo-energy representation learning?

(2) Does explicit geometric information improve prediction relative to a conventional Graph Transformer in a controlled synthetic setting?

(3) How can geometry-aware graph representations support future geo-energy applications such as spatial prediction, geological characterization, connectivity analysis, and decision support?

The main contributions are:

(1) A GAGT framework that incorporates relative geometric information into graph attention for geo-energy systems.

(2) The use of inter-node distance and coordinate displacement as explicit descriptors alongside node attributes and graph connectivity.

(3) A controlled MLP-GT-GAGT comparison that isolates the incremental contributions of graph structure and explicit geometry.

(4) A clear separation between geometry-aware representation learning and physics-informed modeling, together with pathways for combining them.

Future extensions include evaluation across multiple geological realizations and spatial scales, graph construction informed by permeability, flow, or fractures, and integration with caprock, reservoir, and

monitoring information for geological CO_2 and H_2 storage [7, 22].

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents the GAGT formulation. Section 4 describes the experimental design, Section 5 reports the results, and Section 6 discusses implications and limitations. Figure 1 illustrates the problem setting.

2 Related Work and Research Gap

2.1 Data-Driven Learning in Geo-Energy Systems

ML and deep learning have been applied to reservoir characterization, petrophysical prediction, geothermal development, and geological storage [2–5], including reservoir-quality prediction [8, 9], shear-wave-velocity estimation [10, 11, 20], fracture characterization [12–15], graph-based production forecasting, subsurface flow optimization, and CO_2 plume migration prediction [16–19], production indicator prediction in fractured-vuggy reservoirs [21], and storage assessment [22]. These models learn nonlinear relationships among geological, petrophysical, and operational variables effectively. Most of them, however, operate on fixed-dimensional measurements or engineered features that describe local attributes but not the spatial relationships among subsurface locations, which become important when responses depend on spatial propagation, geological boundaries, or the configuration of neighboring regions.

2.2 Graph-Based and Attention-Based Learning

GNNs represent relationships among spatially distributed entities through nodes and edges [23, 24]. In geo-energy applications, nodes can be grid cells, pore-scale regions, wells, or geological subdomains, and edges describe predefined spatial or functional relationships. Pore-GNN, for example, predicts flow properties from pore-scale graph representations [25], and hybrid graph network simulators have been trained to emulate large-scale subsurface flow [31]. Connectivity, however, does not fully characterize geometry: two connected node pairs may differ in distance, displacement, or arrangement despite having the same topological relationship.

The Transformer learns pairwise relationships through attention [26], and graph Transformers extend this mechanism to graph-structured data [27, 28]. Graphormer already adds structural biases, such as shortest-path distance, to the attention score [27], and in 3D point-cloud and molecular learning, relative positions or distances are routinely embedded in attention or message passing [32–34]. In geo-energy graph models, however, spatial geometry usually remains implicit in node features or in the way the graph is built rather than entering the attention score directly.

2.3 Research Gap

Existing geo-energy learning approaches provide feature-based nonlinear prediction, graph-based relational representation, and attention-based interaction learning. What remains insufficiently addressed is the explicit integration of relative geological geometry into graph Transformer attention for subsurface problems. GAGT targets this gap by adding relative geometric descriptors to graph attention while keeping the underlying graph, so that the contributions of node attributes, connectivity, and geometry can be examined separately. GAGT is not a substitute for physics-informed learning [29, 30]; it addresses spatial representation, and geometric proximity is not assumed to represent hydraulic or fracture connectivity.

3 Geometry-Aware Graph Transformer Framework

3.1 Graph Representation

The geo-energy system is represented as an undirected spatial graph $G = (V, E)$, where V is the set of $N = |V|$ spatial nodes and E is the set of spatial connections, with $(i, j) \in E \Rightarrow (j, i) \in E$. Node i carries a feature vector $\mathbf{x}_i \in \mathbb{R}^F$, where F is the number of node features, and the node-feature matrix is

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\top \in \mathbb{R}^{N \times F}. \tag{1}$$

The graph neighborhood of node i is

$$\mathcal{N}(i) = \{\, j \mid (i, j) \in E \,\}, \tag{2}$$

and self-information is retained during aggregation through

$$\mathcal{N}^+(i) = \mathcal{N}(i) \cup \{i\}. \tag{3}$$

The self-loop is used only for aggregation and is not part of E. The overall architecture of GAGT, including node-feature encoding, graph-constrained attention, geometric bias construction, and output prediction, is illustrated in Figure 2.

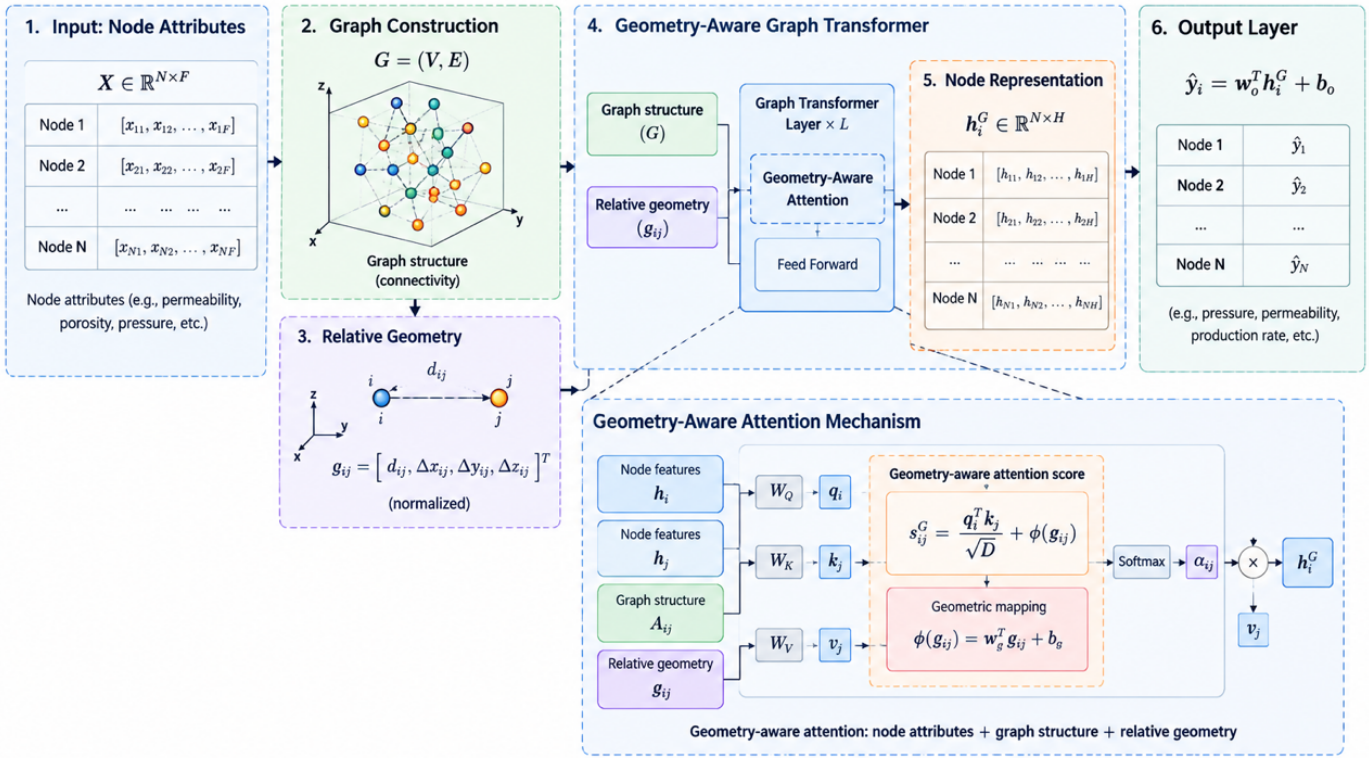

Figure 2. Architecture of the Geometry-Aware Graph Transformer (GAGT). Node attributes, graph structure, and relative geometric information are combined through a learnable geometric bias in the attention mechanism to produce geometry-aware node representations.

3.2 Relative Geometric Representation

Each node i has a spatial coordinate

$$\mathbf{p}_i = [x_i^c,\, y_i^c,\, z_i^c]^{\mathsf{T}}. \tag{4}$$

For a connected pair (i, j), the coordinate differences are

$$\Delta x_{ij} = x_i^c - x_j^c, \quad \Delta y_{ij} = y_i^c - y_j^c, \quad \Delta z_{ij} = z_i^c - z_j^c, \tag{5}$$

and the Euclidean distance is

$$d_{ij} = \sqrt{\Delta x_{ij}^2 + \Delta y_{ij}^2 + \Delta z_{ij}^2}. \tag{6}$$

The raw relative-geometric descriptor is

$$\mathbf{g}_{ij}^{\text{raw}} = [d_{ij},\, \Delta x_{ij},\, \Delta y_{ij},\, \Delta z_{ij}]^{\mathsf{T}}, \tag{7}$$

which is normalized with training-set statistics before model training:

$$\mathbf{g}_{ij} = \text{Norm}\left(\mathbf{g}_{ij}^{\text{raw}}\right). \tag{8}$$

For the self-loop, $\mathbf{g}_{ii} = \mathbf{0}$. The descriptor supplies relative spatial information; it does not assume that geometric proximity is equivalent to hydraulic, fracture, or permeability-based connectivity.

3.3 Graph Transformer Representation Learning

Let $\mathbf{h}_i$ denote the representation of node i at the input of an attention layer (for the first layer, $\mathbf{h}_i$ is obtained from $\mathbf{x}_i$). The query, key, and value vectors are

$$\mathbf{q}_i = \mathbf{W}_q \mathbf{h}_i, \quad \mathbf{k}_j = \mathbf{W}_k \mathbf{h}_j, \quad \mathbf{v}_j = \mathbf{W}_v \mathbf{h}_j, \tag{9}$$

where $\mathbf{W}_q$, $\mathbf{W}_k$, and $\mathbf{W}_v$ are trainable projection matrices. The conventional Graph Transformer attention score between nodes i and j is

$$s_{ij} = \frac{\mathbf{q}_i^{\mathsf{T}} \mathbf{k}_j}{\sqrt{D}}, \tag{10}$$

where D is the dimension of the query and key vectors. The normalized attention coefficient and the updated representation are

$$\alpha_{ij} = \frac{\exp(s_{ij})}{\sum_{l\in\mathcal{N}^+(i)} \exp(s_{il})}, \qquad \mathbf{h}_i' = \sum_{j\in\mathcal{N}^+(i)} \alpha_{ij} \mathbf{v}_j. \tag{11}$$

3.4 Geometry-Aware Attention

GAGT adds relative geometry to the attention score as a learnable bias. The geometric mapping is

$$\psi(\mathbf{g}_{ij}) = \mathbf{w}_g^{\mathsf{T}} \mathbf{g}_{ij} + b_g, \tag{12}$$

with trainable parameters $\mathbf{w}_g \in \mathbb{R}^4$ and $b_g \in \mathbb{R}$. The geometry-aware score, coefficient, and representation are

$$s_{ij}^G = \frac{\mathbf{q}_i^\top \mathbf{k}_j}{\sqrt{D}} + \psi(\mathbf{g}_{ij}), \tag{13}$$

$$\alpha_{ij}^G = \frac{\exp(s_{ij}^G)}{\sum_{l\in\mathcal{N}^+(i)} \exp(s_{il}^G)}, \tag{14}$$

$$\mathbf{h}_i^G = \sum_{j\in\mathcal{N}^+(i)} \alpha_{ij}^G \mathbf{v}_j. \tag{15}$$

The attention weight is thus determined jointly by the learned node representations and the relative geometry of the pair. The formulation does not force shorter distances to receive larger weights; the effect of geometry is learned from data through $\mathbf{w}_g$.

3.5 Relationship to the Conventional Graph Transformer

When the geometric term is removed, $\psi(\mathbf{g}_{ij}) = 0$ and Eq. (13) reduces to Eq. (10), i.e., $s_{ij}^G = s_{ij}$. GAGT therefore contains the conventional Graph Transformer as a special case, which provides a direct basis for the controlled MLP-GT-GAGT comparison.

3.6 Node, Graph, and Geometric Information

The representation separates three sources of information: node features $\mathbf{x}_i$ describe local attributes, the graph G defines admissible spatial relationships, and $\mathbf{g}_{ij}$ describes the relative geometry of connected pairs. Geometry supplements, rather than replaces, the structural information in the graph. This separation matters for geo-energy applications because spatial proximity alone does not necessarily represent effective subsurface interaction; hydraulic properties, fractures, permeability distributions, and boundary conditions may require additional information or domain-informed graph construction. The contribution of each source is assessed by comparing MLP, GT, and GAGT, and the learned geometric weights are examined to see how geometry enters the attention score.

3.7 Output Mapping

For node-level regression, the geometry-aware representation is mapped to the predicted response by a linear output layer:

$$\hat{y}_i = \mathbf{w}_o^\top \mathbf{h}_i^G + b_o, \tag{16}$$

where $\mathbf{w}_o$ and b_o are trainable parameters.

3.8 Extension Toward Physical Consistency

The present framework does not impose governing equations during training. A physical-consistency term could be added in future work through

$$L = L_{\text{data}} + \lambda_{\text{phys}} L_{\text{phys}}, \tag{17}$$

where L_{data} is the prediction loss, λ_{phys} is a weighting coefficient, and, for a generic node-level residual $R_i(\hat{\mathbf{y}})$,

$$L_{\text{phys}} = \frac{1}{N}\sum_{i=1}^{N} \left\| R_i(\hat{\mathbf{y}}) \right\|_2^2. \tag{18}$$

This term is not used in the experiments reported here.

3.9 Summary

GAGT keeps graph connectivity and introduces relative geometry into attention through the trainable bias $\mathbf{w}_g^\top \mathbf{g}_{ij} + b_g$ in Eq. (13). Removing this term recovers the conventional Graph Transformer, which allows the contribution of explicit geometry to be evaluated in a controlled way.

4 Experimental Design

4.1 Controlled Synthetic Spatial Setting

The computational domain contained 300 irregularly distributed nodes in a normalized three-dimensional space, with node positions $\mathbf{p}_i$ defined in Eq. (4). Coordinates were sampled independently within $[0, 1]^3$, which avoids the artificial spatial patterns of a regular lattice. Each node was assigned three local attributes, permeability κ_i, porosity ϕ_i, and pressure P_i, generated independently of the coordinates so that geometry was not implicitly encoded in the node features.

The spatial graph was built with an 8-nearest-neighbor (8-NN) Euclidean rule: an undirected edge $(i, j) \in E$ was created whenever either node was among the other's eight nearest neighbors, with duplicates kept once. The graph contained 1,441 undirected edges, i.e., 2,882 directed message-passing connections. It therefore represents spatial proximity, not hydraulic, fracture, fault, or stratigraphic connectivity. Self-loops were added only through $\mathcal{N}^+(i)$ and were not counted in E; their geometric descriptor was $\mathbf{g}_{ii} = \mathbf{0}$, as in Section 3.

The same node features, graph, targets, data partitions, and preprocessing were used for MLP, GT, and GAGT, so that only the representation differed between models.

4.2 Synthetic Data Generation and Target Construction

The synthetic response was designed so that relative geometry affects neighborhood influence through both distance and direction. Using the differences and distance in Eqs. (5)-(6), a unit direction vector was defined for $d_{ij} > 0$ as

$$\mathbf{u}_{ij} = \frac{[\Delta x_{ij}, \Delta y_{ij}, \Delta z_{ij}]^{\mathsf{T}}}{d_{ij}}, \tag{19}$$

with components $u_{x,ij}$, $u_{y,ij}$, and $u_{z,ij}$.

The geometry-dependent interaction weight was

$$\begin{aligned} w_{ij} = \exp\big(& -d_{ij}/0.20 + 0.55u_{x,ij} \\ & - 0.30u_{y,ij} + 0.20u_{z,ij}\big), \end{aligned} \tag{20}$$

normalized over the neighborhood as

$$\bar{w}_{ij} = \frac{w_{ij}}{\sum_{j' \in \mathcal{N}(i)} w_{ij'}}, \tag{21}$$

where j' is a dummy summation index. The target was then

$$\begin{aligned} y_i = {} & 0.35\, S(\kappa_i) + 0.15\, S(\phi_i) + 0.15\, S(P_i) \\ & + 0.50\, S\Big(\sum_{j \in \mathcal{N}(i)} \bar{w}_{ij} \kappa_j \Big) + 0.35\, S\Big(\sum_{j \in \mathcal{N}(i)} \bar{w}_{ij} P_j \Big) + \epsilon_i, \end{aligned} \tag{22}$$

where ϵ_i is zero-mean Gaussian noise with standard deviation 0.03, and $S(\cdot)$ denotes z-score standardization, $S(x) = (x - \mu)/\sigma$, with μ and σ computed from the training data. The target is a controlled synthetic response, not a governing equation or a physical simulator.

Note that $\bar{w}_{ij}$, built from Eq. (20), is a softmax over the neighborhood of a score that is linear in distance and direction. This has the same functional form as the geometric bias in Eq. (13), so the target mechanism is well matched to the inductive bias of GAGT. The experiment therefore tests whether GAGT can recover a geometry-dependent mechanism when one is present; it does not show that such a mechanism exists in real reservoirs.

4.3 Geometric Representation and Data Normalization

The geometric descriptors of Eqs. (7)-(8) were normalized with statistics computed on the training data only and then fixed for validation and test data. Node attributes were standardized in the same way. $\mathcal{N}(i)$ excludes the self-loop, whereas $\mathcal{N}^{+}(i)$ includes it.

4.4 Model Configurations

The MLP receives $\mathbf{x}_i$ only; GT receives $\mathbf{x}_i$ and G; GAGT receives $\mathbf{x}_i$, G, and $\mathbf{g}_{ij}$. All models used a hidden dimension of 32 and were trained with the Adam optimizer (learning rate 0.003, weight decay 1×10^{-4}) for at most 220 epochs with an early-stopping patience of 35 epochs.

4.5 Data Partition and Repeated Training

The 300 nodes were split into 180 training, 60 validation, and 60 test nodes. The split was fixed and identical for all three models, and the test set was not used for model selection. Ten random seeds, $S = \{11, 22, 33, 44, 55, 66, 77, 88, 99, 111\}$, were used for repeated training; since the data and split were fixed, the seeds varied only the stochastic elements of training (e.g., parameter initialization). All node indices were retained in the experimental package.

4.6 Training Procedure

Each model was trained on the training subset and evaluated on the validation subset after every epoch. The state with the lowest validation loss was retained, and training stopped when the validation loss had not improved for 35 consecutive epochs or after 220 epochs. The retained state was evaluated once on the test subset. Validation-loss histories were saved for all models and seeds.

4.7 Evaluation Metrics

Let T be the set of test indices and $N_{\text{test}} = |T|$. The metrics were

$$\text{MAE} = \frac{1}{N_{\text{test}}} \sum_{i \in T} |y_i - \hat{y}_i|, \tag{23}$$

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i \in T} (y_i - \hat{y}_i)^2}, \tag{24}$$

$$R^2 = 1 - \frac{\sum_{i \in T} (y_i - \hat{y}_i)^2}{\sum_{i \in T} (y_i - \bar{y}_T)^2}, \tag{25}$$

where $\bar{y}_T$ is the mean target over T. The percentage reduction in an error metric M from GT to GAGT is

$$\text{Red}_M = \frac{M_{\text{GT}} - M_{\text{GAGT}}}{M_{\text{GT}}} \times 100\%, \tag{26}$$

and the relative increase in R^2 is computed analogously.

4.8 Statistical Significance and Cross-Seed Variability

The ten seeds were treated as paired runs for the GT-GAGT comparison. The Shapiro-Wilk test was

applied to the paired differences; since normality was not rejected, paired t-tests were used for MAE, RMSE, and R^2. Cross-seed variability of each metric M was quantified by the coefficient of variation

$$\mathrm{CV}_M = \frac{\sigma_M}{|\mu_M|} \times 100\%, \tag{27}$$

where μ_M and σ_M are the mean and standard deviation across the ten runs.

4.9 Quantification of Graph and Geometric Contributions

For MAE and RMSE, the graph and geometric contributions were defined as

$$\Delta M_{\mathrm{graph}} = M_{\mathrm{MLP}} - M_{\mathrm{GT}}, \tag{28}$$
$$\Delta M_{\mathrm{geom}} = M_{\mathrm{GT}} - M_{\mathrm{GAGT}}, \tag{29}$$

and for R^2 as

$$\Delta R^2_{\mathrm{graph}} = R^2_{\mathrm{GT}} - R^2_{\mathrm{MLP}}, \tag{30}$$
$$\Delta R^2_{\mathrm{geom}} = R^2_{\mathrm{GAGT}} - R^2_{\mathrm{GT}}. \tag{31}$$

These differences are model-level comparisons, not a causal decomposition of predictive performance.

4.10 Reproducibility and Scope

The experimental package contains the coordinates, node features, graph edges, targets, data-split indices, normalization statistics, target-generation parameters, random seeds, model configurations, validation histories, test results, statistical tests, and learned geometric parameters. The workflow comprised irregular 3D domain generation, attribute assignment, 8-NN graph construction, relative-geometry computation, target generation, normalization, fixed data partitioning, training, validation-based model selection, test evaluation, and statistical analysis.

5 Results

5.1 Overall Predictive Performance

Table 1 reports the mean test performance and standard deviation over the 10 seeds, and Figure 3 shows the same comparison graphically.

The MLP reached a mean MAE of 0.4652, an RMSE of 0.5754, and an R^2 of only 0.0549. Adding graph connectivity (GT) reduced MAE to 0.2443 and RMSE to 0.3258, i.e., by 47.50% and 43.38%, and raised R^2 to 0.6970. Adding explicit geometry (GAGT) on the same graph and features further reduced MAE to 0.1569 and RMSE to 0.2152, corresponding to reductions of 35.77% and 33.95% relative to GT, while R^2 increased to 0.8678, a relative gain of 24.51%. The GT-to-GAGT improvement thus cannot be attributed to graph connectivity, which both models share.

Table 1. Test performance of the three models across 10 random seeds (mean ± standard deviation).

Model	MAE	RMSE	R^2
MLP	0.4652 ± 0.0150	0.5754 ± 0.0095	0.0549 ± 0.0312
GT	0.2443 ± 0.0037	0.3258 ± 0.0052	0.6970 ± 0.0096
GAGT	0.1569 ± 0.0023	0.2152 ± 0.0019	0.8678 ± 0.0023

5.2 Contribution of Graph Structure and Explicit Geometry

Table 2 summarizes the two successive representation changes defined in Eqs. (28)-(31).

Table 2. Incremental contributions of graph structure and explicit geometry.

Comparison	MAE red.	RMSE red.	ΔR^2 (abs.)
MLP → GT	47.50%	43.38%	0.6421
GT → GAGT	35.77%	33.95%	0.1708

Moving from MLP to GT improved all three metrics substantially, showing the value of modeling interactions among spatially related nodes instead of treating nodes independently. Moving from GT to GAGT reduced MAE by a further 0.0874 and RMSE by 0.1106 and increased R^2 by 0.1708. This is consistent with the target mechanism of Section 4, in which neighborhood influence depends on both distance and direction. It does not imply that shorter Euclidean distance generally means stronger physical connectivity; it indicates that the geometric descriptors allowed GAGT to represent the spatial-response mechanism more accurately in this setting.

5.3 Statistical Significance of the GT-GAGT Difference

The Shapiro-Wilk test did not reject normality of the paired differences in MAE, RMSE, and R^2 ($p = 0.198$, 0.879, and 0.947), so paired t-tests were applied (Table 3).

All three tests gave $p < 0.001$. The negative mean difference for R^2 follows from defining the difference as GT minus GAGT. Because the data split was fixed, these tests show that the GT-GAGT difference is consistent across training runs (initialization and other stochastic elements of training); they do not measure

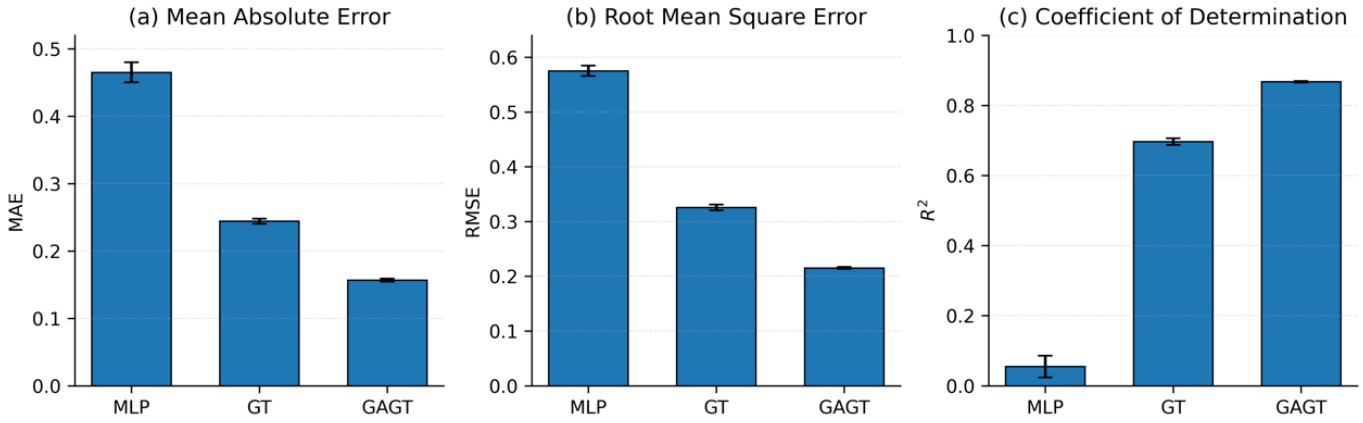

Figure 3. Test-set (a) MAE, (b) RMSE, and (c) R^2 of MLP, Graph Transformer (GT), and Geometry-Aware Graph Transformer (GAGT) over 10 random seeds. Error bars indicate the standard deviation across seeds.

Table 3. Paired statistical comparison between GT and GAGT over 10 seeds.

Metric	Mean diff. (GT−GAGT)	Shapiro-Wilk p	Paired t-test p
MAE	0.0874	0.198	1.43×10^{-12}
RMSE	0.1106	0.879	8.25×10^{-13}
R^2	−0.1708	0.947	2.42×10^{-12}

variability due to different data partitions or geological configurations.

5.4 Cross-Seed Variability

Using Eq. (27), the CVs of GAGT were 1.47% for MAE, 0.88% for RMSE, and 0.26% for R^2, compared with 1.51%, 1.60%, and 1.38% for GT. The MLP showed the largest variation, with CVs of 3.23% for MAE, 1.65% for RMSE, and 56.83% for R^2, the last reflecting its near-zero mean R^2. The low CVs of GAGT indicate that its mean performance is not driven by a few favorable runs; as noted above, they describe sensitivity to training stochasticity under a single synthetic configuration.

5.5 Validation Convergence

Validation-loss histories were recorded for all seeds and models and were used for model-state selection (Section 4). The MLP trajectories varied more across seeds, consistent with its larger cross-seed variation in test MAE. GT converged more stably once neighborhood information was added, and GAGT converged stably to lower validation losses under the same epoch and patience settings. The reported test results come from the validation-selected states rather than the final epoch, so early stopping acted as a common selection procedure rather than a model-specific advantage.

5.6 Learned Geometric Weights

For the representative run with seed 11, the learned weight vector in Eq. (12) was

$$\mathbf{w}_g = [-0.3724,\ 0.3048,\ -0.1535,\ 0.1169]^{\mathsf{T}}, \quad (32)$$

whose components correspond to distance and to the x-, y-, and z-displacements. The negative distance weight lowers the attention score for more distant neighbors, and the directional weights take different signs. The sign pattern $(-,+,-,+)$ matches that of the distance and direction terms in the target-generating weight of Eq. (20) $(-d/0.20,\ +0.55u_x,\ -0.30u_y,\ +0.20u_z)$, which suggests that GAGT used the supplied geometric information to recover the synthetic mechanism. The weights are learned jointly with the node representations and should not be read as physical laws or as evidence that one geological direction is generally more influential; likewise, attention weights are model quantities, not measurements of hydraulic, fracture, or geological connectivity.

5.7 Summary of Findings

Within the controlled experiment, the results provide three pieces of evidence: (i) modeling relationships among connected nodes (MLP→GT) substantially improves prediction over independent node-level learning; (ii) the same graph benefits further from explicit relative geometry (GT→GAGT), with MAE and RMSE reductions of 35.77% and 33.95% and an R^2 gain of 0.1708; and (iii) this improvement is consistent across the ten training runs. These gains are not field-scale evidence. The experiment does not establish whether similar improvements would be obtained with heterogeneous geological structures, irregular

boundaries, faults, fractures, anisotropic permeability, or field well configurations.

6 Discussion

6.1 Contribution of Explicit Geometric Information

The results show a clear progression from MLP to GT to GAGT. Graph connectivity alone does not capture all the spatial information used by the synthetic response mechanism; distance and direction descriptors let GAGT distinguish neighbors by their relative position instead of treating adjacency as a complete description of spatial relationships. The supported conclusion is narrow: explicit relative geometry can add useful information when built into graph attention. It does not follow that Euclidean distance represents hydraulic or geological connectivity.

6.2 Relation to Geo-Energy Spatial Representation

Feature-based models rely on local attributes, and graph-based models add relationships among spatial entities; the present results suggest that these relationships can be refined further by encoding relative geometry. This is relevant to geo-energy systems, where spatial configuration differs considerably among reservoirs, wells, fractures, and geological boundaries. Because the 8-NN Euclidean graph used here encodes proximity only, and no permeability barriers, faults, or stratigraphic interfaces, GAGT in its present form should be regarded as a geometry-aware representation layer rather than a complete geological or physics-informed model.

6.3 Implications for Digital-Twin Decision Support

Geometry-aware graph learning could serve as the predictive layer in digital-twin workflows, supplying spatially structured predictions for tasks such as well-placement screening or injection-rate assessment. This study does not optimize such decisions, so this remains a possible application rather than a demonstrated capability.

6.4 Limitations and Future Research

Four limitations should be noted. First, the evaluation used a synthetic setting rather than field data, so the numerical gains cannot be transferred directly to real geo-energy systems. Second, the target mechanism has the same softmax-over-linear-geometry form as the GAGT attention bias (Section 4); the experiment demonstrates that GAGT can exploit such structure, and benchmarks with simulator-generated or field responses are needed to test it under less favorable conditions. Third, the graph was built by Euclidean 8-NN proximity; permeability-weighted, fracture-informed, and geology-constrained graph constructions should be compared. Fourth, geometry enters only through an attention bias and no physical equations are imposed; a natural extension is to add residual or soft physical constraints through Eq. (17). A practical roadmap is to move from the present controlled experiment to geology-informed graph construction, then to field-scale validation, and finally to physics-constrained, decision-oriented digital-twin applications.

7 Conclusion

This study proposed GAGT, a Graph Transformer that adds a learnable bias based on relative distance and displacement to the attention score. In a controlled synthetic experiment, explicit geometry reduced MAE and RMSE by 35.77% and 33.95% relative to a conventional Graph Transformer using the same graph and node features, and the improvement was consistent and statistically significant across ten training runs. The contribution is methodological: geometry-aware attention can improve graph-based spatial prediction when geometric structure carries information about the response. Validation with field data, simulator-generated responses, and geology-informed graph construction is the next step.

Data Availability Statement

The data generated and analyzed in this study are available from the corresponding author upon reasonable request. The synthetic data can be regenerated from the settings described in Section 4.

Funding

This work was supported without any funding.

Conflicts of Interest

The author declares no conflicts of interest.

AI Use Statement

The author declares that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable. This study did not involve human participants, animals, personal data, or identifiable individual information.

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. [CrossRef]

[2] Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics, 61*, 1-55. [CrossRef]

[3] Zhao, T., Wang, S., Ouyang, C., Chen, M., Liu, C., Zhang, J., ... & Wang, L. (2024). Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innovation, 5*(5), 100691. [CrossRef]

[4] Yan, B., Harp, D. R., Chen, B., Hoteit, H., & Pawar, R. J. (2022). A gradient-based deep neural network model for simulating multiphase flow in porous media. *Journal of Computational Physics, 463*, 111277. [CrossRef]

[5] Okoroafor, E. R., Smith, C. M., Ochie, K. I., Nwosu, C. J., Gudmundsdottir, H., & Aljubran, M. J. (2022). Machine learning in subsurface geothermal energy: Two decades in review. *Geothermics, 102*, 102401. [CrossRef]

[6] Al-Fakih, A., Abdulraheem, A., & Kaka, S. (2024). Application of machine learning and deep learning in geothermal resource development: Trends and perspectives. *Deep Underground Science and Engineering, 3*(3), 286-301. [CrossRef]

[7] Mao, J., & Jahanbani Ghahfarokhi, A. (2024). A review of intelligent decision-making strategy for geological CO_2 storage: Insights from reservoir engineering. *Geoenergy Science and Engineering, 240*, 212951. [CrossRef]

[8] Hu, X., Xie, J., Li, X., Han, J., Zhao, Z., & Ghorbani, H. (2025). Predicting hydrocarbon reservoir quality in deepwater sedimentary systems using sequential deep learning techniques. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources, 11*(1), 96. [CrossRef]

[9] Deng, C., Wang, Y., Mi, W., Xie, X., Sun, X., & Ghorbani, H. (2025). Deep learning-driven analysis of petrophysical dynamics in pay zone quality and reservoir characterization. *Natural Resources Research, 34*(4), 2047-2066. [CrossRef]

[10] Ghorbani, H., Hazbeh, O., Rajabi, M., Tabasi, S., Lajmorak, S., Ahmadi Alvar, M., & Radwan, A. E. (2025). Prediction of formation shear wave velocity using a new hybrid deep learning algorithm. *Physics and Chemistry of the Earth, Parts A/B/C, 140*, 104063. [CrossRef]

[11] Zhao, L., Liu, J., Xu, M., Zhu, Z., Chen, Y., & Geng, J. (2024). Rock-physics-guided machine learning for shear sonic log prediction. *Geophysics, 89*(1), D75-D87. [CrossRef]

[12] Zheng, T., Sun, C., Zhang, J., Ye, J., Rui, X., & Wang, Z. (2024). A multi-aggregator graph neural network for backbone exaction of fracture networks. *Computational Geosciences, 28*(3), 537-549. [CrossRef]

[13] Strait, J. D., Moran, K. R., Hyman, J. D., Viswanathan, H. S., Sweeney, M. R., & Stauffer, P. H. (2023). Fracture network flow prediction with uncertainty using physics-informed graph features. *Computational Geosciences, 27*(6), 1111-1132. [CrossRef]

[14] Srinivasan, G., Hyman, J. D., Osthus, D. A., Moore, B. A., O'Malley, D., Karra, S., ... & Viswanathan, H. S. (2018). Quantifying topological uncertainty in fractured systems using graph theory and machine learning. *Scientific reports, 8*(1), 11665. [CrossRef]

[15] Valera, M., Guo, Z., Kelly, P., Matz, S., Cantu, V. A., Percus, A. G., ... & Viswanathan, H. S. (2018). Machine learning for graph-based representations of three-dimensional discrete fracture networks. *Computational Geosciences, 22*(3), 695-710. [CrossRef]

[16] Liu, W., & Pyrcz, M. J. (2023). Physics-informed graph neural network for spatial-temporal production forecasting. *Geoenergy Science and Engineering, 223*, 211486. [CrossRef]

[17] Li, J., Liu, W., Yu, M., & Xu, W. (2023). Reservoir production prediction based on improved graph attention network. *IEEE Access, 12*, 50044-50056. [CrossRef]

[18] Tang, H., & Durlofsky, L. J. (2024). Graph network surrogate model for subsurface flow optimization. *Journal of Computational Physics, 512*, 113132. [CrossRef]

[19] Ju, X., Hamon, F. P., Wen, G., Kanfar, R., Araya-Polo, M., & Tchelepi, H. A. (2024). Learning CO2 plume migration in faulted reservoirs with Graph Neural Networks. *Computers & Geosciences, 193*, 105711. [CrossRef]

[20] Cova, D., & Liu, Y. (2023). Shear wave velocity prediction using bidirectional recurrent gated graph convolutional network with total information embeddings. *Frontiers in Earth Science, 11*, 1101601. [CrossRef]

[21] Chen, Z., Zhang, D., Li, J., Hui, G., & Zhou, R. (2024). Prediction of production indicators of fractured-vuggy reservoirs based on improved Graph Attention Network. *Engineering Applications of Artificial Intelligence, 129*, 107540. [CrossRef]

[22] Tariq, Z., Ali, M., Yekeen, N., Baban, A., Yan, B., Sun, S., & Hoteit, H. (2023). Enhancing wettability prediction in the presence of organics for hydrogen geo-storage through data-driven machine learning modeling of rock/H_2/brine systems. *Fuel, 354*, 129354. [CrossRef]

[23] Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine, 34*(4), 18-42. [CrossRef]

[24] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A

review of methods and applications. *AI open, 1*, 57-81. [CrossRef]

[25] Alzahrani, M. K., Shapoval, A., Chen, Z., & Rahman, S. S. (2023). Pore-GNN: A graph neural network-based framework for predicting flow properties of porous media from micro-CT images. *Advances in Geo-Energy Research, 10*(1), 39-55. [CrossRef]

[26] 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*. [CrossRef]

[27] Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., ... & Liu, T. Y. (2021). Do transformers really perform badly for graph representation?. *Advances in neural information processing systems, 34*, 28877-28888.

[28] Rampášek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., & Beaini, D. (2022). Recipe for a general, powerful, scalable graph transformer. *Advances in neural information processing systems, 35*, 14501-14515. [CrossRef]

[29] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics, 378*, 686-707. [CrossRef]

[30] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics, 3*(6), 422-440. [CrossRef]

[31] Wu, T., Wang, Q., Zhang, Y., Ying, R., Cao, K., Sosic, R., ... & Leskovec, J. (2022, August). Learning large-scale subsurface simulations with a hybrid graph network simulator. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 4184-4194). [CrossRef]

[32] Fuchs, F., Worrall, D., Fischer, V., & Welling, M. (2020). Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems, 33*, 1970-1981.

[33] Zhao, H., Jiang, L., Jia, J., Torr, P. H., & Koltun, V. (2021). Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 16259-16268). [CrossRef]

[34] Satorras, V. G., Hoogeboom, E., & Welling, M. (2021, July). E (n) equivariant graph neural networks. In *International conference on machine learning* (pp. 9323-9332). PMLR. https://proceedings.mlr.press/v139/satorras21a.html

Jinghan Li is currently a Ph.D. candidate in Management at Lincoln University College, Malaysia. He received the M.Econ. degree from Hiroshima University, Japan, and the B.A. degree from Tiangong University, China. His research interests include artificial intelligence, digital economy, digital twins, sustainable development, environmental risk management, and geo-energy applications. (Email: jinghan.phdscholar@lincoln.edu.my)