Proceedings of the IASS 2024 Symposium
Redefining the Art of Structural Design
August 26-30, 2024, Zurich Switzerland
Philippe Block, Giulia Boller, Catherine DeWolf,
Jacqueline Pauli, Walter Kaufmann (eds.)

Towards Effective Structural Design Embedding by Prompt Engineering Using Large Language Models

Kam-Ming Mark Tam*

[a,]*Department of Architecture, University of Hong Kong
3/F, Knowles Building, The University of Hong Kong, Pokfulam Road, Hong Kong
kmmt@hku.hk

Abstract

This paper investigates the novel application of Large Language Models (LLMs) in structural design as a dimensionality reduction method to embed design descriptions into vector spaces to facilitate design comparison, modelling and generation. Traditional structural design methods, primarily Euclidean-based, struggle to integrate multi-modal descriptions such as text, decision sequences and graphs, particularly when comparing different-sized designs. LLMs emerge as a powerful and versatile method to address this challenge. However, the extent to which LLM embeddings accurately mirror both geometrical and structural design proximities is not well understood. In response to this gap, the research examines how variations in prompt formatting, readability and informativeness impact the effectiveness of LLMs in structural design. The study employs both qualitative and quantitative methods to assess how these prompt characteristics affect the preservation of design attributes and the accuracy of the distances within the embedding space. The findings underscore both the strengths and limitations of LLMs in structural design, offering insights into the capacity of LLMs to facilitate intuitive and effective design comparison and exploration, as well as downstream learning tasks supporting design generation. This work contributes to the effective integration of LLMs into computational structural design frameworks.

Keywords: Structural Design Exploration, Dimensionality Reduction, Machine Learning, Large Language Model, Embedding

1. Introduction

1.1. Large Language Model & Structural Design

In recent years, the emergence of Large Language Models (LLMs) has significantly transformed the AI community [1]. Characterised by their advanced deep learning architectures and extensive training datasets, these models excel in deciphering complex patterns of human language and have demonstrated impressive emergent reasoning capacity [2]. They have extended their utility far beyond their initial remit for automated text generation and natural language, now addressing sophisticated challenges in engineering across multiple domains, including the Architecture, Engineering, and Construction (AEC) sector [3]. Recent attempts in architecture and structural design have explored translating textual descriptions directly into compelling architectural design imageries [4], [5], [6], [7], [8].

Significantly, LLMs offer a viable, unified approach for dimensionality reduction of multi-modal design descriptions—including textual specifications, decision sequences, and graphical representations—into

Copyright © 2024 by Kam-Ming Mark Tam
Published by the International Association for Shell and Spatial Structures (IASS) with permission.

succinct vector embeddings. Traditional Euclidean-based methods falter when analysing designs of varying dimensions and complexities, a limitation LLMs adeptly overcome. Embedding in Machine Learning (ML) generally refers to the transformation of high-dimensional data into a lower-dimensional, dense vector space. In the context of LLMs, this process involves translating detailed design descriptions into text, which is then embedded by LLMs into vectors. These vectors not only facilitate comparisons between designs through measures such as Euclidean distances but also serve as foundational inputs for further learning tasks. ML models can be developed directly from these embedded vectors to predict diverse design attributes or aid in generating new designs [9].

Figure 1 conceptually illustrates how LLMs are used to embed diverse design descriptors into a unified vector space. This process leverages the semantic depth and contextual understanding of LLMs, honed through extensive pre-training on large text corpora, offering potential synergy in structural design comprehension.

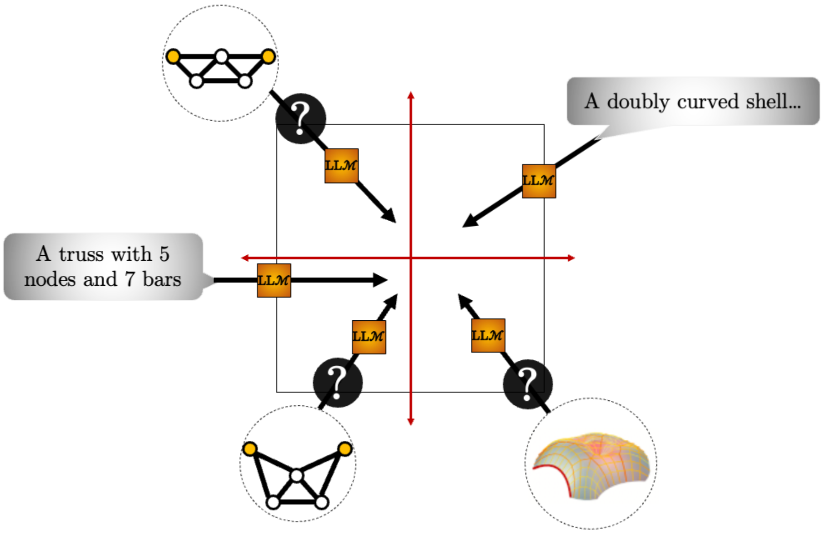

Figure 1: Conceptual depiction of the potential of LLMs to offer an integrated approach for embedding diverse design data descriptions into vector design space raises important questions. However, a critical challenge remains: how should structural designs be described as text?

1.2. Knowledge on Effective Application of Large Language Models in Structural Design

Despite their applicability in design, the use of LLM presents challenges. Crucially, LLMs depend on prompt engineering—the careful design of text inputs—to produce relevant outputs [10], [11]. This process is analogous to feature engineering in traditional ML, where selecting the right attributes is essential for model effectiveness. Anecdotal evidence abounds that minor variations in prompt phrasing can significantly enhance LLM performance for the same tasks, underscoring the model's sensitivity to input configuration.

The integration of LLMs within structural design, particularly in the AEC sector, remains a novel research area. The ability of LLMs to accurately capture and reflect design proximities intuitively is not well understood, highlighting a critical gap that this paper aims to address. Specifically, there is a lack of knowledge regarding optimal translation of structural designs into textual formats that LLMs can effectively process. Furthermore, research on how embeddings from LLMs can be utilised to retrieve structural design information is limited.

This paper aimed to address these gaps by developing insights into how prompt engineering affects the efficacy of LLMs in understanding structural designs. To the author's knowledge, such investigations have not yet been conducted in structural design. Informed by both natural language processing and broader ML research, including investigations on the role of prompt engineering techniques and the analyses of LLM embedding spaces [12], [13], [14], [15], [16], this work tailor a set of methods and protocols to probe the capacity of LLM to understand structural designs. The findings provide a deeper understanding of the models' capabilities and limitations.

2. Methodology

2.1. Investigations

The potential of Large Language Models (LLMs) to effectively embed designs with a nuanced understanding of structural and geometrical qualities was examined via variation on prompt engineering that considered two primary dimensions:

1. **Formatting & Readability:** LLMs are predominantly trained on diverse human-generated text sources such as webpages, conversations, books, and news, with additional inputs from scientific publications and code [10]. The cognitive capacities of these models are, therefore, largely founded on language-based reasoning. This study tested various representations to span a spectrum of human readability, assessing how these affect the model's performance (**§3.2.1**).
2. **Informativeness:** The impact of varying the magnitude of design information input to the LLM on embedding quality was also explored. This was controlled via the number of design features incorporated within the prompts, as detailed in **§3.2.2**.

Furthermore, specific techniques were employed to enhance understanding and control within the embedding process. One approach involved an exhaustive explanation that included a review of fundamental principles in structural design, intended to evaluate how overly descriptive information influences embeddings. Another experiment involved embedding all design attributes discussed in this study to gauge the potential maximum performance of the LLM when comprehensive information is provided. These efforts are organised and summarised in **Table 1**.

Table 1: Dimension of considerations in the design of prompts investigated in this work.

Readability & Formatting (*using* G, S)	'dict' *Dictionary*		(pref., dict.) *Preface + Dictionary*		(pref., sent.) *Preface + Sentence*
Details & Informativeness (*using* (pref., dict.))	G	G, S	G, S, e+	G, S, e+, f+	G, S, e+, f+, g+
Special Considerations	⊛ *with structural principles* (*using* (pref., dict.)+G, S)			g++ *with various global design attributes* (*using* (pref., sent.)+G, S)	

Core design features: G Geometry; S Structure | *Auxillary features*: e+ Edge; f+ Face; g+ Global (see **§3.2.2** & **Table 3**)

2.2. Assumptions & Evaluations

This study investigates the ability of LLMs to effectively embed structural design data, adhering to the principle that a successful embedding should preserve key design attributes and that distances within the embedding space should reflect intuitive distances in actual design contexts. The assessment is structured around several core analyses:

- **Embedding Distribution:** The quality of embeddings is evaluated by examining their distribution and structure using visualisation tools and dimensionality reduction techniques. This analysis helps in understanding the spatial characteristics of the embedded data.
- **Sensitivity to Change:** The study explores how modifications in well-defined design attributes—such as topology, geometry, and internal stiffness—are mirrored in the embedding space. This

involves correlating changes in design features, selected based on their relevance to intuitive and traditional structural design characterisations, with shifts in the embedded representations.

- **Recovery of Information:** This aspect assesses the extent to which specific design attributes can be accurately reconstructed from the embeddings. The ability to recover information from the embedding space provides a measure of the fidelity and utility of the LLM embeddings.
  - Design Groupings: Analyses also include examining how design embeddings align with groups defined by distinct structural characteristics, such as symmetry and types of structural actions.
  - Design Attributes: This includes a detailed evaluation of how various critical geometrical and structural attributes are represented within the embedding space.

2.3. Workflow

The investigation criteria and objectives identified in **§2.2** directly informs the overall structure and workflow of this research, beginning with:

1. **Design Generation:** Designs are created using the pipeline defined in **§3.1**.
2. **Descriptions Generation:** The 9 prompting techniques detailed in **§3.2** are used to generate a textual description for each design.
3. **Text Embedding:** The texts generated in **(2)** are passed to the LLM model specified in **§3.3**, which outputs a vector embedding $\mathbf{z}_{i,p} \in \mathbb{R}^{768}$ that is unique for each design i under a particular prompting technique p.
4. **Dimensionality Reduction:** To improve modelling and reduce computational overhead, each $\mathbf{z}_{i,p}$ is reduced to $\mathbb{R}^{D}$ using Principal Component Analysis (PCA)—a linear dimensionality reduction technique—where D is selected based on a cumulative variance threshold. Specifically, the various analyses make use of two levels of reduced embedding:
   a. **Global:** PCA is applied to all embeddings $\mathbf{z}_{i,p}$ for all techniques at once, resulting in $\mathbf{z}_{\text{PCA}_{\text{Global}},i,p}$.
   b. **Local:** PCA is individually applied to the embeddings $\mathbf{z}_{i,p}$ specific to each technique p, resulting in $\mathbf{z}_{\text{PCA}_{\text{Local}},i,p}$.
5. **Analyses of Embeddings:** All area of investigations introduced in **§2.2** are conducted.

3. Implementation

3.1. Design Dataset

A varied set of N =2000 planar reticulated designs in 2-D was created based on the Author's work in [17]. The process involves sampling-based parameterisation, followed by a two-stage discretisation that includes initialisation followed by an optional subdivision/triangulation application, as **Figure 2** shows.

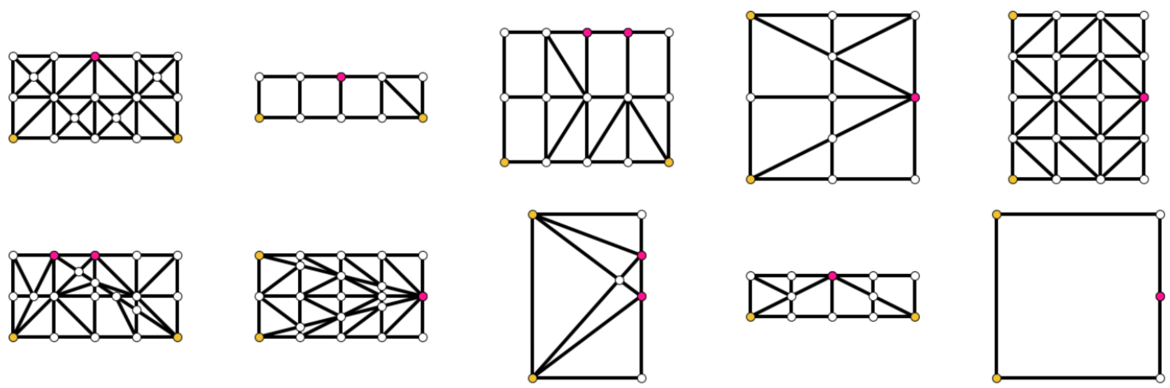

Figure 2: Random selection of 2-D planar reticulated designs of varying degree of structural efficiency embedded in this investigation, indicating the location of supports and loads (*in yellow and pink, respectively*). All interior joints (in *white*) are modeled as rigid.

Moreover, the design generation pipeline was crafted and parameterised to allow the emergence of distinct design characteristics, encapsulating binary differences in loading conditions, connectivity regularity, symmetrical conditions, and more. These qualities contribute to the investigations relying on discrete categorisation in **§4.3**.

3.2. Prompt Engineering

3.2.1. Formatting and Readability

In this work, the concept of formatting and readability were investigated simultaneously in the crafting prompting technique that creates textual descriptions from designs. The strategy directly employs the text corresponding to the Python-based *dictionary* used to describe any given design to develop a prompt that constitutes the least *readable* representation, or (dict.). For improved readiability, (pref.,dict.) prefaces the dictionary with a short description that aims to help contextualise the data it contains as attributes of a 2-D planar reticulated structural design. Finally, (pref.,sent.) expresses a similar preface while replacing the dictionary in its entirety with sentence-based descriptions. Exerpts from all three prompting techniques are displayed in **Table 2**.

Table 2: Prompt engineering corresponding to different degrees of human readability with attributes in G, S.

dict.	`{'vertex': {0:{'x': -4.92, 'y': -1.09, 'p': 0.0, 'is_fixed': False},`
	`1:{'x': -4.92, 'y': 1.09, 'p': 0.0, 'is_fixed': True},`
	`[... more vertices ...],}`
	`'edge': {0:[(0, 1), {'r': 4.72, 'f': 0.0, 'my': [0.0, -0.0], 'l': 2.17, 'SE_ax': 0.0, 'SE_be': 0.0}], 1: [(0, 4), {'r': 4.47, 'f': 0.0, 'my': [0.0, -0.0], 'l': 9.84, 'SE_ax': 0.0, 'SE_be': 0.0}],`
	`[... more edges...]}}`

dict. pref.	`This represents a graph-like 2-D planar structural system. It is encoded as a dictionary with the following format:` `(1) 'vertex': Contains unique integer keys representing vertices, with each vertex specified by a set of attributes including x and y coordinates ('x' and 'y'), support conditions ('is_fixed'), and applied load magnitude ('p'); (2) 'edge': Consists of unique integer keys representing edges, where each edge is defined by a pair of vertex keys encoded as a tuple. Additionally, edge attributes include member lengths ('l'), forces ('f'), bending moments at endpoints ('my'), and the radius of filled circular cross-section ('r'), which is determined according to stress-based criteria, and the strain energies due to axial and bending structural action ('SE_a' and 'SE_b')` `[... same as dict. ..]`
dict. sent.	`This is a graph-like 2-D planar structural system: Vertices are uniquely identified by integer keys and encompass attributes such as x and y coordinates, support conditions (either free or fixed), and the magnitude of applied loads. Edges, distinguished by unique keys, are defined by their boundary vertex keys and are accompanied by attributes such as member forces, endpoint moments, and the radius of a filled circular cross-section, determined based on stress-based criteria.` `Vertices` `- Vertex 0: Coordinates (-2.51, -1.25), Load: 0.0, Support: Fixed,` `- Vertex 1: Coordinates (-1.25, -1.25), Load: 0.0, Support: Free,` `[... more vertices ...]` `Edges:` `- Edge 0: Joins Vertex 0 and 1, Radius: 0.5, Force: -47.85, Moments: [0.03, -0.02],` `- Edge 1: Joins Vertex 0 and 17, Radius: 0.5, Force: -126.58, Moments: [0.02, -0.03],` `[... more edges ...]`
Note:	For (dict.), vertex, edge, and face information are respectively encoded as follows: $\{v_i$: {{key: attr…}} ; $\{e_l$: [(v_j, v_k), {key: attr …}]} and $\{f_m$: [$(v_n, v_{o,}…)$, {key: attr …}]}, where v_i, e_l and f_m denotes respectively integer vertex, edge and face keys; and (v_j, v_k) and $(v_n, v_{o,}…)$ are keys of vertices bounding e_l and f_m, respectively.

3.2.2. Information by Feature Selection

The extent of information embedded is controlled via the selection of features to be incorporated in the construction of prompts for describing designs, which are grouped into five sets of attributes, as shown in **Table 3**. These include (G) and (S), which provide the minimum information required for basic geometrical and spatial understanding, and for describing a 2-D planar reticulated system's structural equilibrium, respectively. Additionally, (e+) and (f+) provides respectively auxillary edge and face attributes, whereas (g+) provides global attributes to summarise a design's overall structural behaviour. Lastly, (g++), which is not detailed in **Table 3**, encapsulates all attributes included in the studies conducted in **§4.3**.

Table 3: Compiled list of vertex, edge, face and global design attributes investigated.

VERTICES	G	S
x-Coordinate	☑	
y-Coordinate		
Is Support		☑
y-Loading		☑

EDGES	S	e^+
Length		☑
Cross Section	☑	
Force	☑	
Moment	☑	
Ax. Strain Energy		☑
Be. Strain Energy		

FACES	e^+	f^+	g^{++}
Area		☑	
GLOBAL			
Strain Energy	☑		
Volume	☑		
(… *various* …)			☑

3.2.3. Attributes to Investigate Informative of Embedding Space

The design features included in the investigations related to the preservation of design information were chosen based on key control parameters in the generative design pipeline, as well as design aspects traditionally considered in the design of reticulated structures. These aspects include their **(A)** topology/connectivity, **(B)** overall geometry, and **(C)** internal stiffness and force distribution. **Figure 3** documents these attributes and their pairwise correlations. Of particular interest are several binary discrete attributes (e.g. the presence of symmetry, diagonals, etc) that are weakly correlated with the other design attributes. This supports their use in **§4.3** as attributes for evaluating the capacity of embeddings to retain design information.

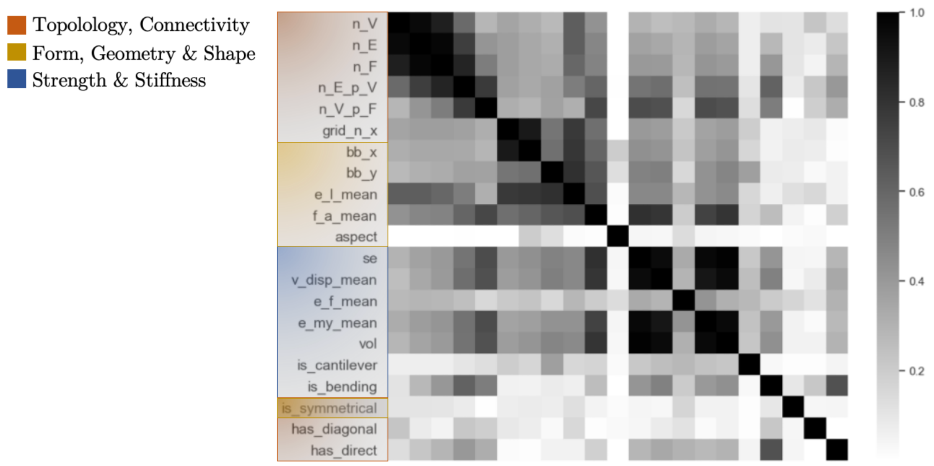

Figure 3: Absolute pairwise Pearson correlation matrix of all attributes studied in the investigations related to design information recovery (i.e. **Table 4**, **Table 5** and **Figure 5**), where each entry $C_{i,j}$ indicates the magnitude of correlation between design aspects i and j, depicted on a color scale ranging from white (0) to black (1).

3.3. Large Language Model

Textual design descriptions are embedded as vector in this work using Longformer [18], which extended Google's Bidirectional Encoder Representations from Transformers [19]. Integrating a sliding window mechanism with attention layers, Longformers can manage larger context and therefore process significantly longer text—up to 4096 tokens—rendering its ideal for handling extensive and verbose structural design descriptions.

4. Investigations & Results

4.1. Visualisation on Embedding Distribution

~~To facilitate direct comparison of the embedding spaces associated to the different prompting techniques, PCA was applied to the entire collection of embedding vectors from all nine techniques, resulting in $\mathbf{z}_{\mathrm{PCA}_{\mathrm{Global}},i,p}$. The first 100 principal components of each embedding vector, capturing 99.8% of the variance of the total embedding space, were then further processed using Uniform Manifold Approximation and Projection (UMAP) [20], a non-linear dimensionality reduction technique that preserves both global and local structure. This two-stage reduction transformed the Longformer-based LLM's complex 768-dimensional embedding space into a more tractable 2-D vector space that can be readily visualised. **Figure 4** displays and compares the transformed embedding spaces for the nine prompting techniques.~~

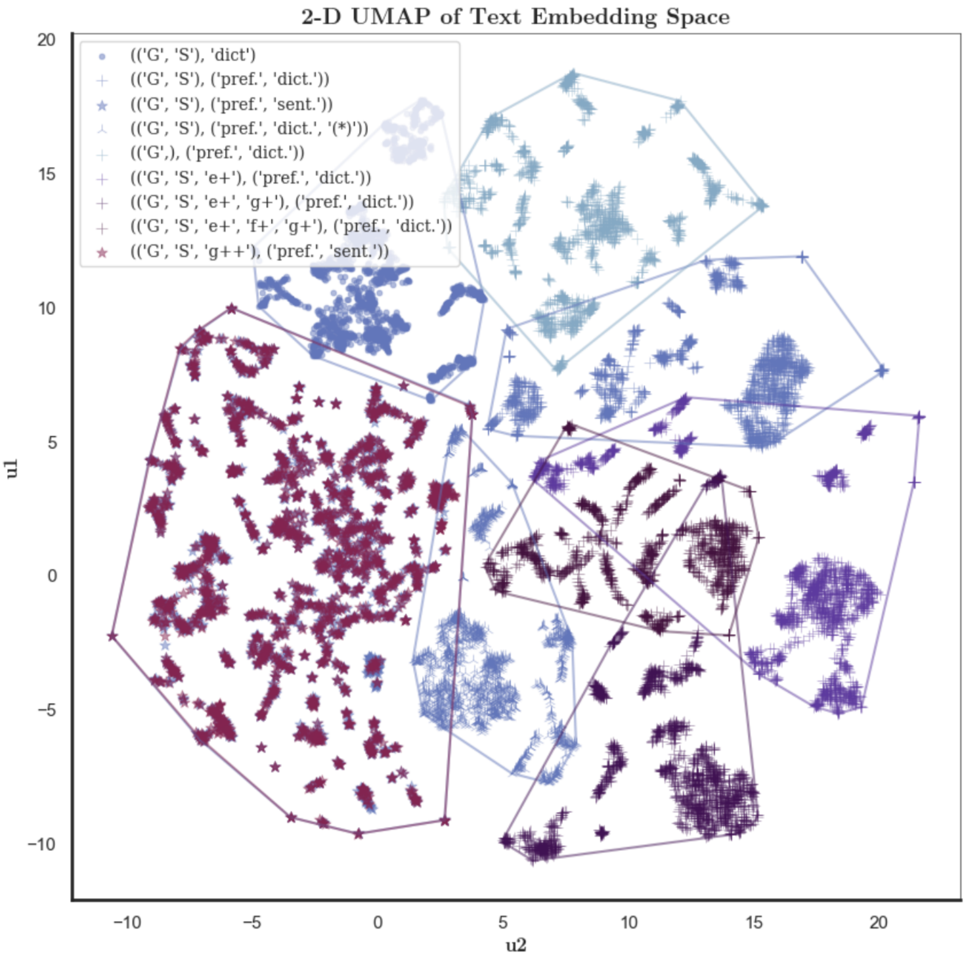

Figure 4: Visualisation of the UMAP-reduced embedding sub-spaces for all design data investigated in this work corresponding to the 9 prompting techniques.

It is evident from **Figure 4** that variations in the readability and formatting of prompts can have a significant impact on how the Longformer positions the design in its embedding space. This is because different prompting techniques occupy different regions of the design space in their embedding. Additionally, the shape and distribution of the embedded space also vary. Visually, the (dict.) approach appears to create a few densely concentrated clusters, while the (pref., dict.) approach spreads the data out more, resulting in more distinct groupings. The effect is most pronounced with the (pref., sent.) approach, where the data is scattered more uniformly with fewer discernible clusters. This suggests a finer differentiation of the underlying design data.

4.2. Sensitivity to Design Change

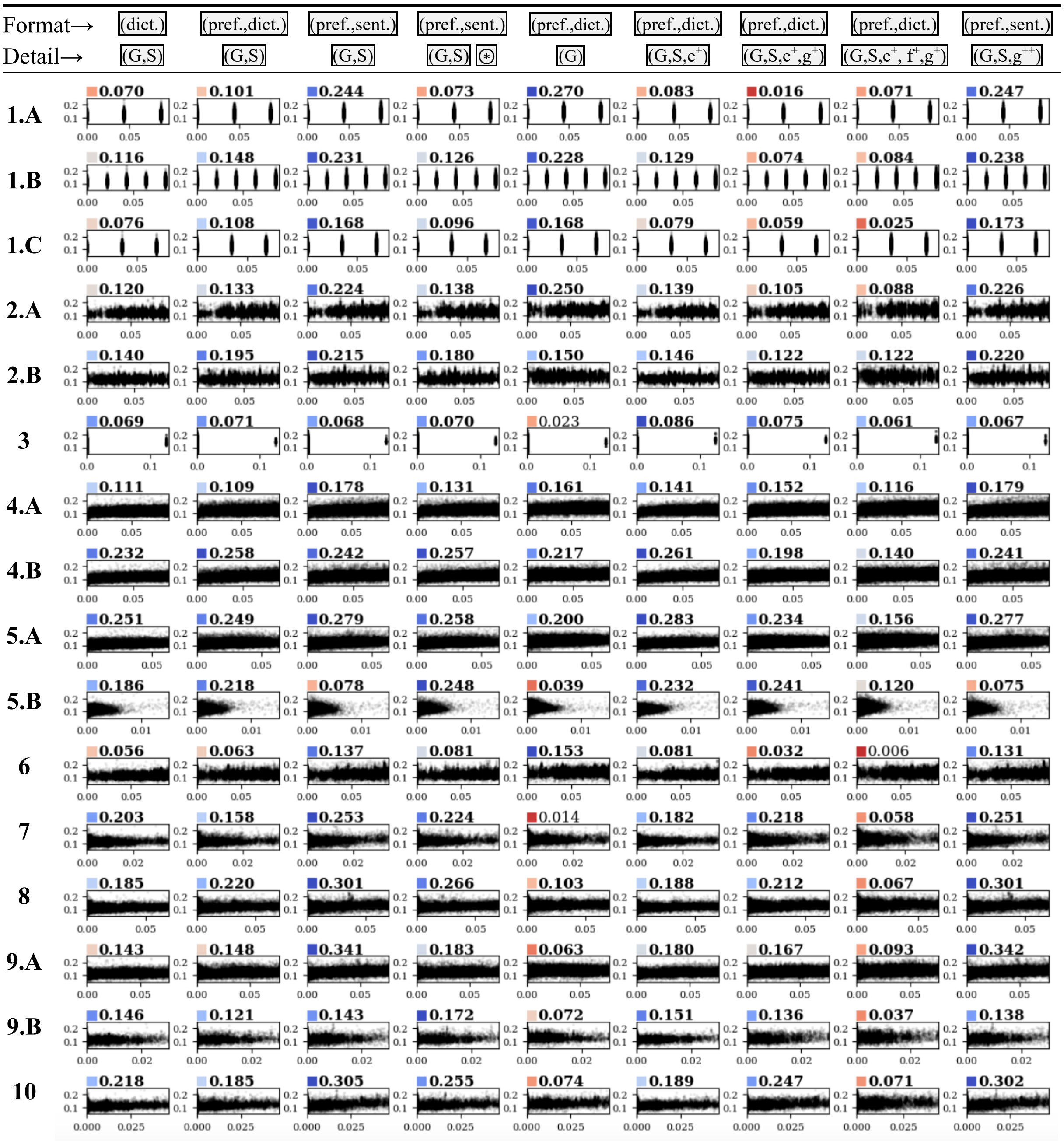

Figure 5: Relating changes in design attributes and corresponding movements in the LLM embedding space under each prompting technique. The Spearman correlation coefficient is displayed above each plot, accompanied by a colour-coded patch indicating its rank across all 9 techniques from worst to best (*red to blue*). Design attributes include: Total number of **(1.A)** Vertices, **(1.B)** Edges, and **(1.C)** Faces. Average number of **(2.A)** Edges per faces, and **(2.B)** Vertices per face; **(3)** Subdivision density along the x-axis; Overall bounding box dimension along the **(4.A)** x- and **(4.B)** y-directions; Average **(5.A)** Edge length and **(5.B)** Face area; **(6)** Aspect ratio of subdivision unit; **(7)** Total strain energy; **(8)** Average vertex displacement; Average magnitude of **(9.A)** Edge force and **(9.B)** Bending moment; and **(10)** Minimum structural volume sized according to stress.

There are indications that different techniques correspond to varying degrees of stability in the embedding of designs. This stability reduces the fluctuations caused by changes in the informativeness of the textual descriptions input to the LLM. For instance, altering the pairing of information with (pref., dict.)–such as using attributes from (G, S, e+), (G, S, e+, f+) or (G, S, e+, f+, g+)–corresponded to significant shifts in the vector embedding. Conversely, pairing (pref., sent.) with information from categories (G, S) or (G, S, g++) results in a similar degree of information change but causes minimal shift in the embedding positions. These results suggest that a more readable sentence-based description of the design provides greater stability, leading to increased consistency and reliability.

Embeddings generated through sentence-based prompting exhibit greater sensitivity to design changes, as evidenced by **Figure 5**. This figure compares the Euclidean distances between approximately 1,000 design pairs (i, j) selected from the dataset across 16 design aspects $x_a \in \{x_1, \dots, x_{A=16}\}$. For each of the 9 prompting techniques investigated, the analysis measured changes in the design aspect under study and their PCA embeddings, represented as $|x_{i,a} - x_{j,a}|$ and $|\mathbf{z}_{\mathrm{PCA}_{\mathrm{Local}}\,i,p} - \mathbf{z}_{\mathrm{PCA}_{\mathrm{Local}}\,j,p}|$, respectively. In this study, the pairs were selected in a controlled manner: each design j paired with a design i had all its attributes within a 10% range of those for i. Furthermore, Spearman's rank correlation coefficients were computed to assess the strength of the monotonic relationships between variables. This method is chosen for its robustness, effectively handling non-linear relations, ordinal data, and not requiring data to conform to a specific distribution. It ranks variables and evaluates their association based on these ranks. Here, only absolute values are shown, with values close to 1 indicating the strongest possible correlation, whether positive or negative. This approach effectively demonstrates how changes in design attributes are reflected in the embedding representations, providing a clear measure of the embedding process's effectiveness.

Several observations align with intuition. For instance, the (pref., dict.)+(G) technique, which only incorporated geometric features, displayed the least sensitivity to structural design attributes (see 7-10 in **Figure 5**). High correlation consistently appeared in prompt techniques involving (pref., sent.). Conversely, the improvement achieved by prefacing raw structural design data encoded as a dictionary (dict.) over (pref., dict.) was marginal, and performance even degraded for certain attributes. Sensitivity also appeared to deteriorate when too much information was provided. This is illustrated by the examples combining (pref., dict.) with (G, S, e+, g++) and (G, S, e+, f+, g+). In fact, the latter case exhibited the second lowest correlations overall. In that vein, (pref., sent.)+(G, S, e+, g++), which directly introduced global attributes corresponding to this study, did not seem to enhance the embedding space's sensitivity to the design attributes when compared to the simpler (pref., sent.)+(G).

4.3. Recovering Design Information

Although proportionate sensitivity may intuitively and numerically seem preferable, it is important to note that it alone is not a sufficient indicator of the embedding space's ability to retain useful information. The relationship between the embedding space and design attributes is likely complex since both the Transformers and Neural Networks that underlie LLM are highly non-linear functional approximators. In this regard, the last two studies employed ML to provide a more systematic and intelligent strategy for assessing the degree of information retained in the latent space. Specifically, **Table 4** and **Table 5** present the accuracies of classifiers and regressors trained to recover respectively key discrete and continuous attributes of designs x_a from their corresponding embeddings $\mathbf{z}_{\mathrm{PCA}_{\mathrm{Local}}\,i,p}$.

In each case, three to four modelling approaches were investigated, and four training repetitions were conducted for each modelling approach. The models were trained using the k-fold cross-validation protocol commonly employed in ML, and the highest accuracy attainable among the different modelling approaches investigated for each attribute was selected and reported as a surrogate measure of the degree of information retained by the embedding space associated with the particular prompting technique.

The classification approaches in **Table 4** include **(A)** Logistic Regression; **(B)** Random Forest; and **(C)** Support Vector Machine with Radial Basis Function kernel. The regression techniques in **Table 5** include **(A)** Linear Regression; **(B)** Epsilon-Support Vector Regression; **(C)** Epsilon-Support Vector Regression with Radial Basis Function kernel; and **(D)** Gradient Boosting. All models were trained until convergence for a maximum of 5,000 iterations.

In general, both results in **Table 4** and **Table 5** reaffirm the observations from **Figure 5** that utilising embeddings from (pref.,sent.) consistently outperformed others, achieving highest accuracies for four-fifths of the discrete attributes and the lowest errors for 13 out of 16 continuous attributes. Notably, these models reached near-perfect accuracy in classifying designs based on their load cases, such as cantilever versus simply supported structures.

Table 4: Table displaying the highest average accuracy across three training repetitions for three binary classifier models using embeddings obtained from the nine prompting strategies. The classifiers were evaluated on five design categories: **(1)** Problem case (cantilever or simply supported); **(2)** Symmetry; presence of **(3)** bending behaviour; **(4)** triangulation, and **(5)** direct load-support connection. Cell colouring indicate the rank in model accuracy across all 9 techniques from worst to best (*red to blue*).

	(dict.)	(pref.,dict.)	(pref.,sent.)	(pref.,sent.)	(pref.,dict.)	(pref.,dict.)	(pref.,dict.)	(pref.,dict.)	(pref.,sent.)
	(G,S)	(G,S)	(G,S)	(G,S) ⊙	(G)	(G,S,e^{+})	(G,S,e^{+},g^{+})	(G,S,e^{+}, f^{+},g^{+})	(G,S,g^{++})
1	0.97±1.2e-2	0.96±8.6e-3	**0.99±1.6e-3**	0.97±9.0e-3	0.92±1.1e-2	0.97±5.0e-3	0.97±7.8e-3	0.94±6.4e-3	0.99±2.3e-3
2	**0.65±9.7e-3**	0.63±1.5e-2	**0.65±1.3e-2**	0.62±1.0e-2	0.62±9.7e-3	0.62±9.8e-3	0.61±9.5e-3	0.61±7.6e-3	0.64±8.4e-3
3	0.84±6.3e-3	0.77±1.0e-2	**0.85±6.5e-3**	0.78±1.0e-2	0.75±4.5e-3	0.78±1.0e-2	0.77±6.9e-3	0.74±1.3e-2	0.85±2.7e-3
4	**0.74±6.6e-3**	0.73±5.3e-3	**0.74±7.2e-3**	0.73±6.6e-3	**0.74±5.6e-3**	0.73±5.4e-3	0.73±1.0e-2	**0.74±5.4e-3**	**0.74±6.9e-3**
5	0.80±1.3e-2	0.75±7.8e-3	**0.82±9.7e-3**	0.75±1.3e-2	0.81±9.0e-3	0.77±7.1e-3	0.76±1.2e-2	0.78±6.8e-3	0.83±9.0e-3

Table 5: Table displaying the minimum average error across three training repetitions for three regressors using embeddings obtained from the nine prompting strategies. The regressors were evaluated under Mean Absolute Percentage Error on their prediction of various design attributes: Total number of **(1.A)** Vertices, **(1.B)** Edges, and **(1.C)** Faces. Average number of **(2.A)** Edges per faces, and **(2.B)** Vertices per face; **(3)** Subdivision density along the *x*-axis; Overall bounding box dimension along the **(4.A)** *x*- and (4.B) *y*-directions; Average **(5.A)** Edge length and **(5.B)** Face area; **(6)** Aspect ratio of subdivision unit; **(7)** Total strain energy; **(8)** Average vertex displacement; Average magnitude of **(9.A)** Edge force and **(9.B)** Bending moment; and **(10)** Minimum structural volume sized according to stress. Cell colouring indicate the rank in model error across all 9 techniques from worse to best (*red to blue*).

	(dict.)	(pref.,dict.)	(pref.,sent.)	(pref.,sent.)	(pref.,dict.)	(pref.,dict.)	(pref.,dict.)	(pref.,dict.)	(pref.,sent.)
	(G,S)	(G,S)	(G,S)	(G,S) ⊙	(G)	(G,S,e^{+})	(G,S,e^{+},g^{+})	(G,S,e^{+}, f^{+},g^{+})	(G,S,g^{++})
1.A	0.07 ± 6.7e-4	0.09 ± 2.1e-3	0.05 ± 4.6e-4	0.10 ± 1.6e-3	0.06 ± 6.8e-4	0.10 ± 8.1e-4	0.10 ± 1.4e-3	0.09 ± 7.3e-4	**0.04 ± 4.5e-4**
1.B	0.13 ± 1.2e-3	0.17 ± 2.3e-3	0.12 ± 1.0e-3	0.19 ± 2.4e-3	0.13 ± 1.5e-3	0.18 ± 2.2e-4	0.18 ± 3.2e-3	0.16 ± 2.5e-3	**0.11 ± 4.7e-4**
1.C	0.28 ± 4.3e-3	0.34 ± 7.6e-3	**0.26 ± 2.4e-3**	0.35 ± 4.8e-3	0.27 ± 5.7e-4	0.34 ± 3.8e-3	0.34 ± 5.3e-3	0.33 ± 5.1e-3	0.27 ± 2.5e-3
2.A	0.09 ± 1.1e-3	0.11 ± 4.1e-4	**0.09 ± 1.1e-4**	0.10 ± 5.7e-4	0.09 ± 4.2e-4	0.11 ± 3.4e-4	0.11 ± 7.5e-4	0.1 ± 3.0e-4	**0.09 ± 2.4e-4**
2.B	0.11 ± 9.3e-4	0.12 ± 5.3e-4	**0.10 ± 1.9e-4**	0.12 ± 1.2e-3	0.10 ± 7.4e-4	0.12 ± 1.8e-4	0.12 ± 1.0e-3	0.11 ± 6.1e-4	**0.10 ± 4.3e-4**
3	0.10 ± 1.1e-3	0.10 ± 8.4e-4	**0.09 ± 6.8e-4**	0.09 ± 3.0e-4	0.11 ± 1.3e-3	0.09 ± 1.1e-3	0.09 ± 9.0e-4	0.11 ± 1.4e-3	0.09 ± 1.9e-3
4.A	0.13 ± 1.8e-3	0.14 ± 2.7e-3	0.15 ± 1.3e-3	0.13 ± 1.2e-3	0.14 ± 4.3e-4	0.12 ± 1.3e-3	**0.11 ± 1.4e-3**	0.13 ± 2.2e-3	0.15 ± 5.6e-4
4.B	0.23 ± 2.7e-3	0.21 ± 3.1e-3	**0.18 ± 8.2e-4**	0.22 ± 2.2e-3	0.19 ± 4.2e-4	0.23 ± 2.0e-3	0.23 ± 1.6e-3	0.26 ± 3.6e-3	0.19 ± 1.6e-3
5.A	0.14 ± 3.1e-4	0.15 ± 2.1e-3	0.14 ± 1.6e-3	0.14 ± 2.3e-3	0.16 ± 9.5e-4	0.13 ± 1.0e-3	0.14 ± 3.4e-3	0.16 ± 3.0e-3	0.15 ± 6.0e-4
5.B	**0.45 ± 1.5e-3**	0.48 ± 5.2e-3	**0.45 ± 3.3e-3**	0.52 ± 1.1e-2	0.53 ± 1.2e-2	0.48 ± 7.8e-3	0.49 ± 6.9e-3	0.52 ± 7.6e-3	0.48 ± 3.7e-3
6	0.12 ± 1.6e-3	0.12 ± 1.2e-3	**0.11 ± 9.6e-4**	0.12 ± 5.8e-4	**0.11 ± 1.7e-4**	0.12 ± 2.9e-4	0.12 ± 6.5e-4	0.12 ± 6.4e-4	**0.11 ± 1.3e-3**
7	0.76 ± 3.6e-3	0.82 ± 1.2e-2	**0.67 ± 3.2e-3**	0.84 ± 1.1e-2	0.86 ± 5.3e-3	0.80 ± 6.6e-3	0.82 ± 6.8e-3	0.85 ± 1.9e-3	**0.67 ± 5.2e-3**
8	0.77 ± 4.7e-3	0.83 ± 1.6e-2	**0.69 ± 4.8e-3**	0.84 ± 8.2e-3	0.86 ± 5.5e-3	0.80 ± 4.8e-3	0.83 ± 9.7e-3	0.86 ± 1.8e-3	**0.69 ± 8.7e-3**
9.A	0.33 ± 8.7e-4	0.32 ± 5.0e-4	**0.28 ± 2.1e-3**	0.33 ± 4.2e-3	0.36 ± 2.3e-3	0.33 ± 1.0e-3	0.33 ± 1.4e-3	0.35 ± 7.2e-4	**0.28 ± 1.9e-3**
9.B	1.52 ± 3.2e-3	1.61 ± 9.4e-4	**1.50 ± 4.9e-3**	1.59 ± 3.0e-3	1.64 ± 6.6e-3	1.59 ± 5.0e-3	1.60 ± 3.2e-3	1.60 ± 5.0e-3	1.51 ± 1.3e-3
10	1.08 ± 4.8e-3	1.17 ± 4.6e-3	**1.00 ± 4.6e-3**	1.18 ± 4.9e-3	1.20 ± 2.5e-3	1.14 ± 6.6e-3	1.16 ± 4.6e-3	1.20 ± 2.6e-3	**1.00 ± 4.3e-3**

However, significant errors were observed in the regressors, particularly concerning attributes related to average bending magnitude and minimal structural volume requirements determined by stress criteria. These discrepancies likely arise from the non-linear scaling disproportion of these attributes compared to others, compounded by the minimal model parameter fine-tuning conducted, which is computationally intensive and outside this paper's scope. Nevertheless, these areas are anticipated to show improvement with additional fine-tuning efforts.

5. Discussion

Several key insights can be extracted from this study. Firstly, the investigation supports the idea that LLMs are well-suited for language-based reasoning. The notable improvement achieved by (pref.,sent.) over (pref.,dict.), which performed less effectively compared to dict., suggests a potential need for consistent formatting and representation in encoding design information—either adopting entirely sentence-based or code-like dictionary formats.

Additionally, the quality of the embedding space depended on the quantity of embedded information. The findings indicate that adding extra attributes does not always lead to a more informative representation. While including structural information (S) benefited the comparison of embedding spaces associated with information set (G), adding information on face connectivity and attributes seemed to have a crowding-out effect, resulting in models trained on these embeddings exhibiting lower overall accuracy. Moreover, providing the model directly with the attributes under investigation had a negligible effect on the preservation of these attributes in the embedding spaces. These nuances underscore the importance of optimising the feature engineering of information to be embedded as prompts in any application of LLMs.

6. Conclusion

6.1. Limitations & Outlook

This study, while focusing on the efficacy of Longformer-based Large Language Models (LLMs) in structural design embeddings, highlights several areas for future research:

- Model Generalisability: The distinct architectures and training datasets of LLMs suggest that behaviours observed here may differ with other models. Future studies should explore whether the insights on model efficacy, human readability, and language-based intelligence are consistent across alternative LLM architectures.
- Scope of Design Systems: Current findings are limited to reticulated truss or frame structures and predominantly 2-D designs. Expanding this research to include a broader variety of design systems is crucial for generalising these results.
- Model Optimisation: This research employed both linear and non-linear models in classification and regression to evaluate information preservation within embedding spaces. Further refinement and fine-tuning are necessary to maximise data extraction, possibly incorporating neural network-based approaches.
- Handling Larger Designs: The limited token length of LLMs poses challenges for encoding more complex designs with numerous nodes and connections. Developing strategies for effectively abstracting and describing large-scale systems remains an essential area for future exploration.

6.2. Summary

The presented research advances the application of LLMs in embedding structural design descriptions, underscoring their utility in enhancing design comparison and generation. The study establishes a comprehensive protocol with methods tailored for evaluating the effectiveness of LLM embeddings within structural design contexts. These methods facilitated the refinement of prompting techniques, with findings indicating that uniform sentence-based formatting significantly boosts model performance. Moreover, while the inclusion of pertinent structural attributes typically enriches the embeddings, an excess of (irrelevant) details can impair performance, highlighting the critical role of optimally calibrating feature engineering. These insights lay the groundwork for more strategic integration of language processing technologies into computational structural design frameworks.

Acknowledgements

The author expresses sincere appreciation to Prof. Dr. Pierluigi D'Acunto and Prof. Dr. Karla Saldaña Ochoa for inspiring this study on the application of LLM in structural design, and to Prof. Dr. Eric H.

Schuldenfrei for a conversation on prompt engineering that provided important motivations for this piece. Special thanks to Lazlo Bleker for insightful discussions. Heartfelt gratitude also goes to the Block Research Group, especially Prof. Dr. Philippe Block and Dr. Tom Van Mele, for their enduring support and significant contributions to the author's growth and development.

References

[1] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2009.

[2] J. Wei *et al.*, "Emergent Abilities of Large Language Models," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.07682

[3] A. Saka *et al.*, "GPT models in construction industry: Opportunities, limitations, and a use case validation," *Developments in the Built Environment*, vol. 17. Elsevier Ltd, Mar. 01, 2024. doi: 10.1016/j.dibe.2023.100300.

[4] C. Bannon, "Carlos Bannon." 2024.

[5] Z. Zhang, J. M. Fort, and L. Giménez Mateu, "Decoding emotional responses to AI-generated architectural imagery," *Front Psychol*, vol. 15, p. 1348083, 2024.

[6] A. Fernández-Morales, "The Possibilities of Text-to-Image Tools for the Generation of Floor Plans," in *Congreso Internacional de Expresión Gráfica Arquitectónica*, 2024, pp. 297–307.

[7] N. O. Hanafy, "Artificial intelligence's effects on design process creativity:" A study on used AI Text-to-Image in architecture"," *Journal of Building Engineering*, vol. 80, p. 107999, 2023.

[8] E. Yildirim, "Text-to-image generation AI in architecture," *Art and architecture: theory, practice and experience*, vol. 97, 2022.

[9] Z. Guo, K. S. Ochoa, P. D. ' Acunto, K. Saldana Ochoa, and P. D'acunto, "Enhancing structural form-finding through a text-based AI engine coupled with computational graphic statics." [Online]. Available: https://www.researchgate.net/publication/363845839

[10] W. X. Zhao *et al.*, "A Survey of Large Language Models," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.18223

[11] G. Marvin, N. Hellen, D. Jjingo, and J. Nakatumba-Nabende, "Prompt Engineering in Large Language Models," 2024, pp. 387–402. doi: 10.1007/978-981-99-7962-2_30.

[12] I. Keraghel, S. Morbieu, and M. Nadif, "Beyond words: a comparative analysis of LLM embeddings for effective clustering." [Online]. Available: https://hal.science/hal-04488175

[13] Z. Nie, R. Zhang, and Z. Wu, "A Text is Worth Several Tokens: Text Embedding from LLMs Secretly Aligns Well with The Key Tokens," Jun. 2024.

[14] K. Edemacu and X. Wu, "Privacy Preserving Prompt Engineering: A Survey," Apr. 2024, [Online]. Available: http://arxiv.org/abs/2404.06001

[15] A. Petukhova, J. P. Matos-Carvalho, and N. Fachada, "Text clustering with LLM embeddings," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.15112

[16] A. Bhargava, C. Witkowski, M. Shah, and M. Thomson, "What's the Magic Word? A Control Theory of LLM Prompting," Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.04444

[17] K.-M. M. Tam, "Learning Discrete Equilibrium: Trans-topological Inverse Pattern and Force Design Using Machine Learning and Automatic Differentiation," Zurich, 2023.

[18] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The Long-Document Transformer," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.05150

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805

[20] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform Manifold Approximation and Projection," *J Open Source Softw*, vol. 3, no. 29, p. 861, Sep. 2018, doi: 10.21105/joss.00861.