LLM-to-Phy3D: Physically Conform Online 3D Object Generation with LLMs

Melvin Wong[1], Yueming Lyu[2], Thiago Rios[3], Stefan Menzel[3], Yew Soon Ong[1,2]

[1]*College of Computing & Data Science (CCDS), Nanyang Technological University (NTU), Singapore*
[2]*Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR), Singapore*
[3]*Honda Research Institute Europe (HRI-EU), Offenbach am Main, Germany*

{wong1357, asysong}@ntu.edu.sg, lyu_yueming@a-star.edu.sg, {thiago.rios, stefan.menzel}@honda-ri.de

***Abstract*—The emergence of generative artificial intelligence (GenAI) and large language models (LLMs) has revolutionized the landscape of digital content creation in different modalities. However, its potential use in Physical AI for engineering design, where producing physically viable artifacts is paramount, remains largely underexplored. The lack of physical knowledge in existing LLM-to-3D models often results in outputs that are detached from real-world physical constraints. To address this gap, we introduce LLM-to-Phy3D, an online 3D object generation method that enables existing LLM-to-3D models to produce physically conforming 3D objects on the fly. LLM-to-Phy3D introduces a novel online black-box refinement loop that empowers large language models (LLMs) through synergistic visual and physics-based evaluations. By delivering directional feedback in an iterative refinement process, LLM-to-Phy3D actively drives the discovery of prompts that yield 3D artifacts with enhanced physical performance and geometric novelty bias relative to user reference objects, marking a substantial contribution to AI-driven generative design. Systematic evaluations of LLM-to-Phy3D, supported by ablation studies in vehicle design optimization, reveal various LLM improvements gained over 78% in producing physically conform target domain 3D designs over conventional LLM-to-3D models. The encouraging results suggest the potential general use of LLM-to-Phy3D in Physical AI for scientific and engineering applications.**

***Index Terms*—prompt evolution, vision-language model, large language models, generative AI, text-to-3D, evolutionary optimization, engineering design optimization**

I. Introduction

Remarkable advancements in generative artificial intelligence (GenAI) and large language models (LLMs) have opened new research opportunities, primarily focused on creating high-fidelity digital artifacts that align with user preferences through tailored prompts. This has led to a surge in broad applications, including text-to-*X* generation, which produces digital artifacts across different modalities (e.g., text, images, or 3D) conditioned on textual prompts [1]–[3].

Among those applications, text-to-3D object generation has attracted increasing attention due to its success and potential to democratize 3D content creation, accelerate design workflows, and leverage the capabilities of emerging multimodal foundation models [3]. One major challenge in 3D object generation is automating the design process while synthesizing novel and diverse shapes [4]. Several works introduce text-to-3D models that can generate the parameters of implicit functions for rendering textured meshes and neural radiance fields [5], [6]. This enables the synthesis of diverse 3D shapes in a target domain from a single text prompt. Xiang et al. [7] proposed a unified structured latent representation that expands the generation of outputs to the 3D Gaussian format and improves the quality of generated artifacts. Wang et al. [8] attempted to unify spatial textual knowledge in LLMs with 3D information to improve mesh understanding, thereby enabling the generation of simple polygonal meshes from text.

While these advancements led to the proliferation of digital media applications, there is a notable lack of progress in advancing 3D object generation in Physical AI for engineering design, where the production of physically viable artifacts is paramount [9], [10]. One major challenge we observed is generating objects that satisfy physical constraints. 3D vehicles generated from conventional text-to-3D models without physical knowledge (Fig. 1 (Left Group) and Fig. 2 (a)) while aesthetically pleasing, have limited efficacy in both the real physical world and its digital twin as compared to ones generated with these constraints (Fig. 1 (Right Group) and Fig. 2 (b)). Notable work attempts to incorporate such constraints into GenAI models, focusing on enforcing the generation of a target object shape that complies with physical laws [11]. However, due to confidentiality concerns and prohibitive data-generation costs, the scarcity of large-scale, high-quality engineering data hinders the scaling and generalization of such techniques for text-to-3D generation. In contrast, black-box techniques can overcome such difficulties, enabling existing GenAI models to generate novel artifacts that conform to targeted physical constraints. Pioneering research studies that investigate this approach in engineering design optimization, where preliminary results, particularly in enhancing aerodynamic performance while satisfying domain constraints [12], [13], encourage progress and interest. These efforts motivate the use of GenAI with language models to narrow the gap between digital innovation and functional applicability, driving progress in Physical AI for engineering design. However, these methods focus on improving conventional optimization strategies. To date, the feasibility of using LLMs to generate 3D artifacts that adhere to physical laws in engineering design remains largely unexplored.

To this end, we introduce *LLM-to-Phy3D*, a novel physically-

Fig. 1. Most representative vehicles generated with various conventional LLM-to-3D models using LLM-generated prompts (left) and physically conforming novel vehicles generated with LLM-to-Phy3D (right).

conforming online 3D object generation method with LLMs. Central to LLM-to-Phy3D is its unique online black-box iterative refinement framework that adaptively steers the LLM to find textual prompts that yield physically plausible 3D artifacts with high geometric novelty. The framework starts with 3D artifacts generated with randomly sampled LLM-generated prompts. Subsequently, LLM-to-Phy3D quantifies the physical performance, target-domain alignment, and geometric novelty of 3D artifacts using visual and physics-based evaluations, which serve as surrogate indicators for a non-differentiable black-box objective. An iteration of LLM-to-Phy3D ends with a selection process that picks the best-performing LLM-generated prompts as exemplars, enabling the LLM to learn the relationship between its prompts and the generated 3D artifacts that yielded the quantified score, thereby improving its performance in the next iteration. This selection pressure mechanism not only ensures constant space complexity but, more importantly, achieves the desired generation characteristics with directional feedback [14], [15] that steers the LLM to produce physically conforming target domain 3D artifacts.

Moreover, generation issues, such as hallucinations [16], produce 3D artifacts whose shapes are outside the target domain yet still satisfy the physical constraints. Visual evaluations can not only effectively identify and penalize irrelevant shapes but also reward novel ones. While existing foundation vision models can assess such artifacts from 2D rendered views with high accuracy [17], [18], projecting 3D artifacts onto 2D can introduce significant geometric distortions in surface topology, degrading model performance. To address such issues, we propose conducting the physical light simulation under orthographic projection [19] to capture the precise surface topology of 3D artifacts and minimize geometric distortions. This single visual representation serves the dual purpose of object recognition and quantification of geometric novelty relative to reference objects.

To demonstrate the efficacy of LLM-to-Phy3D against conventional LLM-to-3D methods, we conduct experiments on a test scenario that mimics real-world engineering design optimization. Starting from initial LLM-generated prompts that yield 3D artifacts with significant variation in physical performance, we demonstrate that all LLMs in our experiments are able to improve the quality of its generated prompts, leading to significant improvements in producing physically conforming novel 3D artifacts with better realism than the ones generated with conventional text-to-3D models. To summarize, the contributions of this paper are as follows:

- We introduce *LLM-to-Phy3D*, a novel physically conform online 3D object generation with LLMs, for actively steering the LLM to produce physically conforming novel 3D objects. LLM-to-Phy3D enables the LLM to find effective prompts by providing physical knowledge through directional feedback. This allows the LLM to learn the association between its generated prompts and the quality of 3D artifacts generated with its these prompts.
- LLM-to-Phy3D captures the precise surface topology and visual characteristics of 3D artifacts through physical light simulation with orthographic projection in a single visual representation for object recognition and geometric novelty quantification with respect to reference objects. The technique minimizes geometric distortion, ensuring the rendered surface topology conforms to scientific and engineering requirements.
- Visual and physics-based evaluations are introduced to quantify the quality of 3D-generated artifacts based on its physical performance and geometric novelty. These evaluations contribute to the search objective, enabling the selection of physically conforming ones as exemplars for the LLM to learn from in the next iteration.
- Systematic evaluations of LLM-to-Phy3D, along with ablation studies in aerodynamic vehicle design optimization tasks, reveal various LLM improvements gained over 78% in producing physically conforming target domain 3D designs, indicating the broader potential of the framework in Physical AI for science and engineering applications.

II. Preliminaries

In this paper, we articulate the physical viability of 3D objects in the context of, but not limited to, automotive aerodynamic design. As such, we introduce in this section the key fundamental concepts in representing and quantifying the novelty and quality of 3D objects.

Surface Topology Representation for Object Recognition and Geometric Novelty Quantification: In many computer vision and computer graphics applications, the semantic meaning and visual features of a 3D object are defined based on the object's surface topology [20]–[22]. This topology information

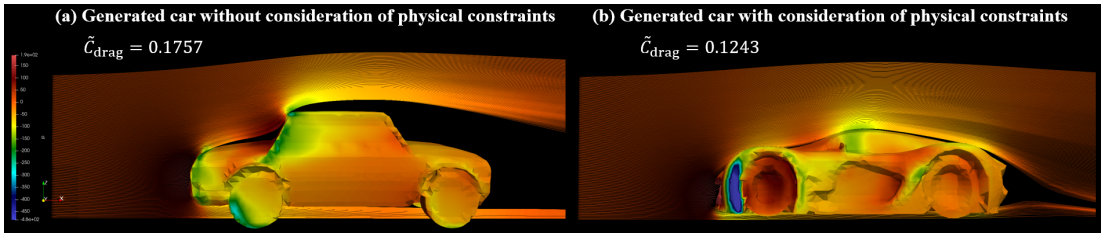


Fig. 2. Aerodynamic Visualization of two vehicles generated (a) without considering physical constraints and (b) with physical constraints. Vehicle shape has a direct impact on the aerodynamic (drag) performance. Some aerodynamically efficient (low-drag) characteristics include attached, streamline flow from the front to the rear and small pressure differences between the front and rear. In contrast, an aerodynamically inefficient (high-drag) vehicle exhibits the opposite characteristics, including separated flow and pronounced pressure differences.

can be captured by illuminating the surface with a physical light simulation [19], which replicates light rays reflecting and refracting off the surface and projecting onto a camera plane. Assuming camera orthographic projection, given a set of all intersection points $\mathbf{Z}_c$ in a fixed camera viewpoint c of a 3D-object $\mathbf{x}$, the surface topology of the rendered R multi-view is defined as:

$$g(\mathbf{x}) = \{\mathbf{X} = \{I(\mathbf{z}) \,|\, \forall \mathbf{z} \in \mathbf{Z}_c\} \,|\, \forall c \in \{1, ..., R\}\}, \quad (1)$$

where $\mathbf{X}$ is a rendered view image of $\mathbf{x}$ consisting of the illumination (or shading) from multiple light rays intersecting at different surface point $\mathbf{z}$ by $I(\cdot)$. In this paper, we compare the multi-view of generated and reference 3D objects to minimize inaccuracies in determining the region of geometric novelty due to surface topology distortion caused by indirect illumination [23].

Aerodynamics Computation for Physical Performance Quantification: For automotive aerodynamics [24], the drag and lift characteristics of a vehicle are of prime interest. These characteristics can be quantified using a physics simulator (or surrogate) G in which the vehicle is subjected to fluid-flow conditions, such as steady or unsteady air streams, thereby replicating real-world aerodynamic environments. The fluid particles interacting with the vehicle's surface generate viscous forces and pressure in the flow direction that drag or lift the vehicle body. Assuming the particles are travelling along the x-axis, given the projected area of the vehicle ($A_{\text{projected}}$), the drag and lift scores are defined as:

$$\begin{aligned} G_{\text{fluid\_forces}}(\mathbf{x}, \mathcal{F}) &\approx \frac{\mathcal{F}(\mathbf{x})}{2\rho_\infty \|\mathbf{u}_\infty\|_2^2 \cdot A_{\text{projected}}}, \\ \bar{C}_{\text{lift}}(\mathbf{x}) &= G_{\text{fluid\_forces}}(\mathbf{x}, \mathcal{F}_z), \\ \bar{C}_{\text{drag}}(\mathbf{x}) &= G_{\text{fluid\_forces}}(\mathbf{x}, \mathcal{F}_x), \end{aligned} \quad (2)$$

where ρ_∞ is the freestream ambient density of the fluid, and $\mathbf{u}_\infty$ is the freestream velocity vector. Here, $\mathcal{F}(\cdot)$ computes the integral forces acting on the surface with $\mathcal{F}_x(\cdot)$ considering the forces along the x-axis only while $\mathcal{F}_z(\cdot)$ take into account forces along the z-axis only. Note that the drag and lift coefficients ($\bar{C}_{\text{drag}}$ and $\bar{C}_{\text{lift}}$) are usually computed in a dimensionless and normalized form.

III. Method: LLM-to-Phy3D

We first present the LLM-to-Phy3D framework, followed by detailed discussions of its key components.

A. Online Blackbox Iterative Refinement Framework

Without loss of generality, let $\mathcal{P}$ denote a space of textual prompts and $\mathcal{X}$ denote a space of 3D objects. Given any 3D object $\mathbf{x} \in \mathcal{X}$, we assume a text-to-3D GenAI model (p_{t3d}) that generate new 3D objects conditioned on a textual prompt. Moreover, given any prompt $\mathbf{m} \in \mathcal{P}$, we also assume the presence of an LLM-based optimizer (p_{llm}) that can iteratively generate new prompts based on the domain of interest $\mathcal{S}$ (e.g., "car", "watercraft", etc) and a meta-prompt $\mathbb{M}$, and can be adaptively guided using exemplars consisting of prompts paired with its performance in generating high-quality 3D objects. As such, given a set $\mathbf{Y}$ of reference 3D objects, the search goal is to find novel 3D artifacts that satisfy physical and domain constraints:

$$\begin{aligned} \min_{\mathbf{m} \in \mathcal{P}} : \; & \mathop{\mathbb{E}}_{p_{\text{t3d}}(\mathbf{x} \,|\, \mathbf{m})} [f_{\text{physical}}(\mathbf{x}) + f_{\text{domain}}(g(\mathbf{x}), \mathcal{S}) - \beta F_{\text{novelty}}(\mathbf{x}, \mathbf{Y})] \\ \text{s.t.} \;\; & Q(\mathbf{m}, \mathcal{S}) \leq \epsilon, \\ & \underline{c} \leq f_{\text{physical}}(\mathbf{x}) \leq \bar{c}, \end{aligned} \quad (3)$$

where $Q(\cdot, \cdot)$ is a function that filters infeasible LLM-generated prompts, due to generation failure scenarios (e.g. LLM hallucination), which are semantically distant to the domain of interest $\mathcal{S}$ by ϵ. Moreover, $\underline{c} > 0.0$ and $\bar{c} < 1.0$ define the boundary that any 3D generated artifact in the target domain can realistically achieve in the real world, effectively filtering infeasible ones such as the non-conforming or non-watertight with object's surfaces are not fully closed. In addition, d ensures the search does not deviate far from the target domain and get stuck in infeasible regions. The first term ranks artifacts with better physical performance higher than those with impractical performance. Among those that satisfy the physical constraint, generated artifacts less resembling $\mathcal{S}$ are penalized in the second term, demoting its position for such artifacts. Finally, the third term rewards artifacts with geometric novelty bias relative to the user reference objects, thereby promoting their positions. This term is weighted with a β parameter to balance the novelty score with the physical score. These three terms work in synergy to provide the LLM with effective feedback on the quality of generated artifacts, ensuring that higher-quality candidates serve as exemplars for subsequent iterations.

Algorithm 1 describes the mechanics of the framework. Given a domain of interest $\mathcal{S}$, LLM-to-Phy3D initializes the LLM (p_{llm}) with the search task. Our proposed method then samples prompts from the LLM to produce the corresponding

3D artifacts from the text-to-3D generative model p_{t3d}. The generated 3D artifacts are evaluated against the specifications to determine the scores. These scores are then combined with the LLM-generated prompts to serve as exemplars for the LLM to learn and re-attempt to create new prompts that can yield better-performing candidates. These candidates are evaluated and, together with the ones generated in the last iteration, undergo a selection process. This process selects the better-performing candidate from each random pair to form an exemplar set of N candidates for the next iteration (Line 7 in Algorithm 1), thereby ensuring constant space complexity throughout the search. Moreover, the exemplar set serves as directional feedback, allowing it to learn the association between the prompt and the quality of the 3D object and find more effective prompts that steer towards the optimal. LLM-to-Phy3D iterates until the terminating condition is satisfied.

Algorithm 1 LLM-to-Phy3D

Input: $\mathbb{M}$ meta prompt, $\mathcal{S}$ domain of interest, batch-size N
Output: Designs Set $\mathcal{D}_{\text{step}}$

1: step $\leftarrow 0$
2: $\mathcal{D}_{\text{step}} \leftarrow \{(\mathbf{m}_i, \mathbf{x}_i) \,|\, \mathbf{m}_i \sim p_{\text{llm}}(\mathbf{m}_i \,|\, \mathbb{M}, \mathcal{S}),\ \mathbf{x}_i \sim p_{\text{t3d}}(\mathbf{x}_i \,|\, \mathbf{m}_i)\} \forall i \in \{1, ..., N\}$
3: Evaluate $\mathcal{D}_{\text{step}}$ against $\mathcal{S}$ based on Equation (3), (4), (5), and (7)
4: **while** non-terminating condition **do**
5: $\quad \mathcal{D}_{\text{candidates}} \leftarrow \{(\mathbf{m}_i, \mathbf{x}_i) \,|\, \mathbf{m}_i \sim p_{\text{llm}}(\mathbf{m}_i \,|\, \mathbb{M}, \mathcal{S}, \mathcal{D}_{\text{step}}),\ \mathbf{x}_i \sim p_{\text{t3d}}(\mathbf{x}_i \,|\, \mathbf{m}_i)\} \quad \forall i \in \{1, ..., N\}$
6: $\quad$ Evaluate $\mathcal{D}_{\text{candidates}}$ against $\mathcal{S}$ based on Equation (3), (4), (5), and (7)
7: $\quad \mathcal{D}_{\text{step+1}} \leftarrow$ select_N_candidates$(\mathcal{D}_{\text{candidates}} \cup \mathcal{D}_{\text{step}})$
8: $\quad$ step $\leftarrow$ step $+ 1$
9: **end while**

B. Quality of 3D Generated Objects

We quantify the quality of a 3D-generated object $\mathbf{x}$ with the following measures:

(I) Physical Constraint with Physical Alignment Measure: This measure assesses the performance of 3D-generated objects under various physics conditions, governed by the natural physical laws. Here, we provide an instantiation of this measure in the context of fluid dynamics [24]. As such, the physical alignment measure is defined as follows:

$$f_{\text{physical}}(\mathbf{x}, \mathcal{F}) = \left[\min(\max(G(\mathbf{x}, \mathcal{F}), a), b) - a\right] / (b - a), \quad (4)$$

where a physics simulator (or surrogate) $G(\cdot, \cdot)$, such as a Computational Fluid Dynamics (CFD) solver [25], computes the physical performance of the 3D generated artifact $\mathbf{x}$.

(II) Domain Constraint with Domain Alignment Measure: Given a domain of interest $\mathcal{S}$ and multi-view images $g(\mathbf{x})$ of the 3D object $\mathbf{x}$, we measure the degree of alignment between the object and target domain with the following cosine similarity measure:

$$\begin{aligned} s_1 &= H_{\text{vlm}}(g(\mathbf{x}), \mathcal{S}), \\ s_2 &= H_{\text{vlm}}(g(\mathbf{x}), \neg\mathcal{S}), \\ f_{\text{domain}}(g(\mathbf{x}), \mathcal{S}, \Gamma) &= e^{s_1/\Gamma} / (e^{s_1/\Gamma} + e^{s_2/\Gamma}), \end{aligned} \quad (5)$$

where Γ is the temperature hyperparameter that controls the discriminative strength between the positive and negative pairs (e.g., $\mathcal{S}$ = "car" and $\neg\mathcal{S}$ = "No car"). Here, we assess the likelihood that a 3D object is within the domain of interest $\mathcal{S}$ with a Visual-Language Model $H_{\text{vlm}}(\cdot, \cdot)$ that sets $\Gamma = 0.01$ [17].

(III) Geometric Novelty Measure: This measure detects and localizes the regions of geometric novelty by comparing the surface topology of a 3D-generated object with a reference object. The surface topology differences with the reference object indicate regions of geometric novelty bias that contribute to the improvement or degradation of the overall physical performance f_{physical}. Here, we instantiate this measure in the context of physical light simulation [19] to capture the precise surface topology of a 3D object. As such, given a 3D-generated object $\mathbf{x}$ and a target object $\mathbf{y}$ in the conventional set, the measure compares images of the same rendered view, defined as follows:

$$\begin{aligned} s_3 &= \sum_{i=1}^{R} \|\mathbf{M} \odot g(\mathbf{y})^{[i]} - \mathbf{M} \odot g(\mathbf{x})^{[i]}\|_2^2, \\ s_4 &= \sum_{i=1}^{R} \sum_{j=1}^{J} \|h_j(\mathbf{M} \odot g(\mathbf{y})^{[i]}) - h_j(\mathbf{M} \odot g(\mathbf{x})^{[i]})\|_2^2, \\ f_{\text{novelty}}(\mathbf{x}, \mathbf{y}) &= \frac{\gamma(g(\mathbf{x}), g(\mathbf{y}))(s_3 + s_4)}{\|\mathbf{M}\|_1}, \end{aligned} \quad (6)$$

where $\odot$ denotes the element-wise product, R is the number of rendered views, $g(\cdot)^{[i]}$ provides the i-th rendered view of a 3D object, and $\mathbf{M}$ is a mask matrix for indicating the regions of interest to compare the surface topology. The first term s_3 compares the same rendered view image between the 3D-generated and target object at the pixel level. The second term s_4 compares various visual feature maps with a pre-trained image feature extractor $h_j(.)$, which extracts specific features at the j-th level of image abstraction. On a separate note, the measure score is weighted with $\gamma(\cdot, \cdot)$ to ensure the relevance of the 3D-generated novelty with respect to the target domain. This allows impractical 3D-generated objects that have no resemblance to the target domain to be ranked lower than the practical ones. In addition, the f_{novelty} is scale invariant by normalizing with the number of pixels considered in the regions of interest $\|\mathbf{M}\|_1$. Furthermore, if a set $\mathbf{Y}$ of reference objects is provided, the reference object with the least geometric novelty with respect to $\mathbf{x}$ is selected as the representative novelty of the set and is defined as:

$$F_{\text{novelty}}(\mathbf{x}, \mathbf{Y}) = \min_{k \in \{1, ..., K\}} \{f_{\text{novelty}}(\mathbf{x}, \mathbf{y}_k) \,|\, \mathbf{y}_k \in \mathbf{Y}\}. \quad (7)$$

TABLE I
COMPARISON OF LLM-TO-PHY3D AND BASELINES *DPAR* (MEAN $\pm$ STD EQ. (8)) IN PRODUCING PHYSICALLY CONFORM TARGET DOMAIN 3D ARTIFACTS

Text-to-3D Model / LLM	**Shap-E** [6]		**Trellis** [7]	
	Baseline	**Ours**	**Baseline**	**Ours**
Mistral 3.0 Small	0.0868 ± 0.0247	$\mathbf{0.9347} \pm 0.0263$	0.0831 ± 0.0576	$\mathbf{0.8412} \pm 0.0460$
GPT 3.5	0.1123 ± 0.0092	$\underline{\mathbf{0.9655}} \pm 0.0271$	0.4167 ± 0.0209	$\underline{\mathbf{0.9578}} \pm 0.0322$
GPT 4o Mini	0.0927 ± 0.0068	$\underline{\mathbf{0.9558}} \pm 0.0462$	0.2618 ± 0.0048	$\underline{\mathbf{0.9768}} \pm 0.0267$
Gemini 2.0 Lite	0.0401 ± 0.0055	$\mathbf{0.9125} \pm 0.0048$	0.2449 ± 0.0021	$\mathbf{0.9472} \pm 0.0106$
Overall	0.0830 ± 0.0299	$\mathbf{0.9059} \pm 0.0362$	0.2517 ± 0.1221	$\mathbf{0.9308} \pm 0.0615$

C. Meta Prompt Design

A set of instructions (or meta-prompt) is designed for the LLM to function as an effective optimizer. Specifically, a role description highlights the search task the LLM operates and the goal (or objective) it needs to solve. LLM is provided with target-domain knowledge to generate domain-specific prompts, effectively constraining the generation of 3D objects within that domain. To this end, we instruct the LLM to complete the following prompt template, $\mathcal{M}$ = "A $< \mathcal{S} >$ in the shape of" , where $\mathcal{S}$ is the domain of interest. Lastly, for the LLM to improve its generative performance, we provide the LLM with feedback by presenting previously LLM-generated prompts with its evaluated scores as exemplars for the LLM to reflect on and learn from. This in-context learning approach improves the likelihood that the LLM generates more effective prompts than its previous iteration.

IV. EXPERIMENTS

We demonstrate the efficacy of LLM-to-Phy3D through experiments designed to solve an engineering design optimization problem. In this section, we will first present the experiment setup, followed by the evaluation, and finally the results and discussion.

A. Setup

The problem of interest is to discover novel cars with aerodynamically efficient (low drag) body shapes. As such, the target domain $\mathcal{S}$ is set as "*Car*" and we design the meta prompt for the problem with the search goal of minimizing the aerodynamic drag $\bar{C}_{drag}$ in f_{physical}. In addition, we bound the drag scores within $a = -1.0$ and $b = 1.0$ in Equation (4) and normalized. Moreover, we set $d \leq 0.7$ in Equation (3). Furthermore, we set β according to only reward novel cars with better aerodynamic drag performance than a set of typical car objects $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_Q\}$. Specifically, we first sample reference car objects in $\mathbf{Y}$ with the text-to-3D generative model in the test scope conditioned on "*A car*", and then determine the physical performance of each object. Subsequently, we compute the physical performance standard deviation $\hat{\sigma}$ and mean $\hat{\mu}$ of the set $\mathbf{Y}$ for instantiating the hyperparameter as $\beta = e^{-(\hat{\mu}/\hat{\sigma})}$. We employ CFD physics simulation with OpenFoam [26] to obtain the physical performance aerodynamic scores. In addition, to generate the feature maps in Equation (6), we utilize the EfficientNet model [18]. On a separate note, we employ the Nvidia ray tracing engine OPTIX [19] to capture and render the precise surface topology of 3D objects.

Our proposed method is compared against four different baseline LLM-to-3D methods. We select four LLMs - *GPT-3.5 Turbo*, *GPT-4o Mini*, *Gemini 2.0 Flash Lite*, and *Mistral 3.0 Small*, to compare based on its capabilities in following the designed meta-prompts and creating feasible prompts throughout the iterations in a single experiment run. As these LLMs require a text-to-3D generative model to produce 3D artifacts, we select *Shap-E* [6] and *Trellis* [7] models. For these baseline LLM-to-3D methods, we established a standard meta prompt "*Create a prompt that starts with $\mathcal{M}$ and ends with a full stop.*", where $\mathcal{M}$ is the prompt template introduced in Section III-C. Moreover, the temperature of the LLMs is set in a range that allows it to produce viable and diverse prompts. Also, for consistency, we use the same temperature range for these LLMs, unless notable hallucination issues significantly degrade the LLM's ability to follow instructions.

In all experiments, we use the same random seeds across multiple runs, with each run producing $N = 20$ candidates generated per iteration. We ran each experiment for 40 optimization steps, yielding up to 800 designs per run. Furthermore, in a single experiment run, the experiment is performed on a single shared compute node with a configuration comprised of Intel Xeon Silver 64 CPU cores clocked at 2.10 GHz, 128GB of RAM, and three Nvidia Quadro GV100 GPUs (32 GB each).

B. Evaluation

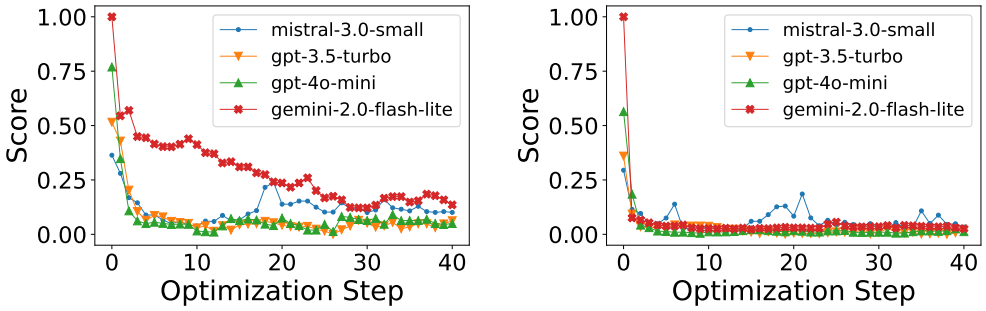
We compare LLM-to-Phy3D against baseline methods in terms of its ability to produce physically conforming target-domain 3D artifacts. Specifically, given the target domain $\mathcal{S}$ and set $\boldsymbol{\mathcal{D}}$ of all 3D artifacts generated with a method in a single experiment run, we define the domain and physical alignment rating (*DPAR*) of a method of interest as:

$$\textit{DPAR}(\boldsymbol{\mathcal{D}}, \mathcal{S}) = \mathop{\mathbb{E}}_{\mathbf{x} \in \mathcal{D}} \left[f_{\text{domain}}\left(g(\mathbf{x}), \mathcal{S}\right) / f_{\text{physical}}(\mathbf{x}) \right], \quad (8)$$

where the higher the rating, the better the method at producing physically conforming target-domain artifacts. For a fair comparison, we ensure that the number of generated outputs is the same for the baseline and our proposed methods across iterations. Furthermore, our proposed method and the baseline use the same LLM temperature ranges.

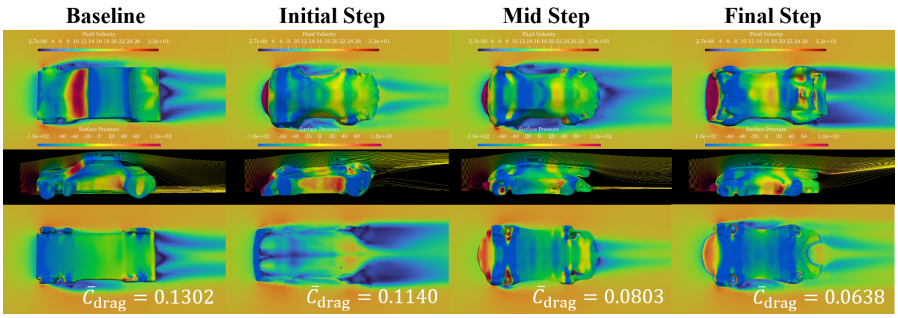
C. Results and Discussions

Based on the experimental results, the baseline models often generate vehicles with aesthetically pleasing shapes or that resemble conventional vehicle body shapes (as observed in

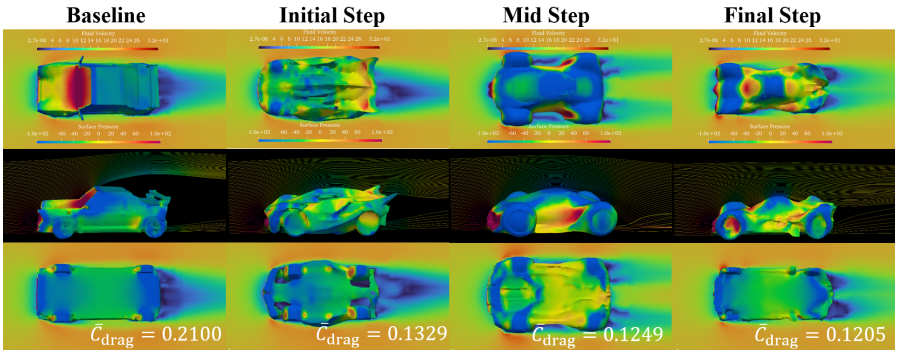


(a) Shap-E [6] **(b) Trellis** [7]

Fig. 3. Search optimization score (Equation 3) of LLM-to-Phy3D with different LLM and text-to-3D generative models combinations.



(a) Shap-E



(b) Trellis

Fig. 4. Results of the aerodynamic profile of cars generated with LLM-to-Phy3D at specific iterations and compared with a representative car generated with the best baseline LLM-to-3D method. Note that the lower the drag, the better the car's physical performance.

some of the most representative examples in the left group of Fig. 1). In contrast to these vehicles, LLM-to-Phy3D found novel vehicles that are not only visually creative but also satisfy the physical and target-domain constraints (as shown in the right group of the same figure). This observation is evident in the benchmark with *DPAR* (Table I), where our proposed method consistently outperforms the baselines across different LLMs of interest by more than 78%. LLM-to-Phy3D demonstrates its efficacy in enabling each LLM to learn the association between 3D artifact quality and its LLM-generated prompts, and in predicting the best possible prompts that yield higher-quality outputs.

As presented in Fig. 3, *GPT-4o-mini* and *GPT-3.5-Turbo* can learn more proficiently than other LLMs of interest, resulting in convergence in less than 10 steps. On the other hand, while *Gemini 2.0 Flash Lite* is capable of producing diverse and creative prompts, we observe it requires more time to overcome the generation issues in Shap-E (Fig. 3(a)) where the text-to-3D model is less adhering to the LLM-generated prompts. We also noticed that *Mistral 3.0 small* struggles to learn the association between its generated prompts and the evaluation scores, hindering its ability to produce effective textual prompts. While still able to find physically plausible and more novel 3D vehicles, the limitations observed in *Gemini 2.0 Flash Lite* and *Mistral 3.0 small* resulted in minimal variation in the generated prompts.

By analyzing the aerodynamic performance of the generated 3D shapes (Fig. 4), our findings reveal that the initialization step produces 3D-generated cars with shapes that yield high surface pressure at the car front and sides. This increases the aerodynamic drag force and decreases the performance. These resulting artifacts have similar aerodynamic issues to the baseline samples, as the LLMs generate textual prompts without prior knowledge of the physics. After introducing prior

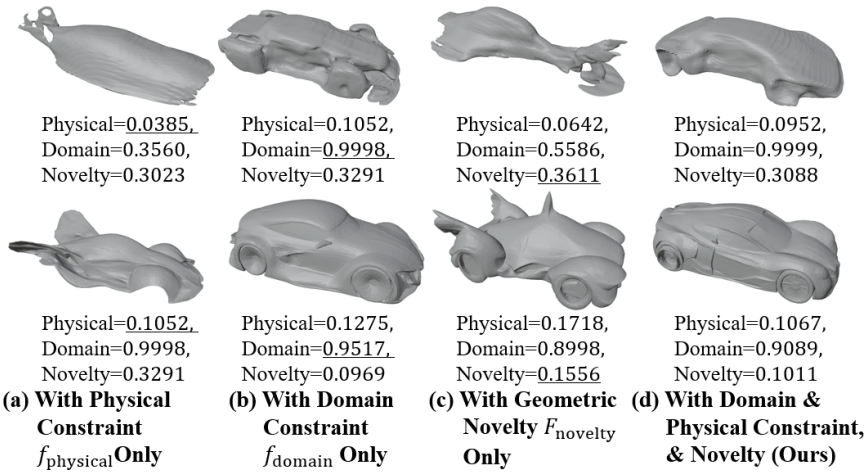


Fig. 5. Ablation Studies on the effectiveness of different terms in the objective function on generating 3D artifacts (Top Row: LLM with Shap-E, Bottom Row: LLM with Trellis).

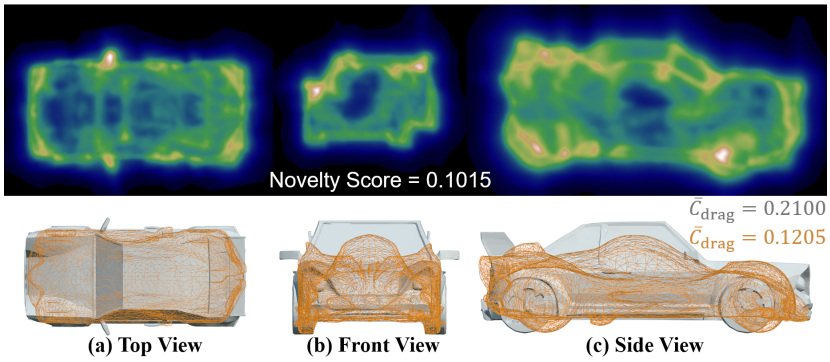


Fig. 6. Example of geometric novelty between a reference car (in grey solids) and a car generated with LLM-to-Phy3D (in orange wireframe). The novelty heatmaps presented in the top row are the multi-view results of $F_{\text{novelty}}(.)$, which highlight the regions contributing to the geometric novelty score of the generated car compared with the target one. Multiple views (bottom row) of the generated car overlay on the target car show regions corresponding to contrasting body shapes that contribute to the difference in aerodynamic performance.

knowledge and selecting effective textual prompts that result in better physically-conforming 3D artifacts, the generated 3D artifacts possess shapes that lead to lower aerodynamic performance indicators, such as a reduced pressure difference or smaller projected area.

We conduct several ablation studies to examine the effect of each term in the objective function. We noticed that using physical constraint f_{physical} only, the search favors malformed or highly fragmented cars (Fig. 5 (a)) that registered better drag performance than complete generated cars. This degrades the framework's overall performance over time as the LLM learns false positives, leading to incorrect associations of prompts with cars and resulting in shapes that perform worse in drag. On a separate note, the search process adhering to the target domain constraint f_{domain} only can lead to artifacts that visually highly resemble the reference cars of varying physical performance (Fig. 5 (b)). With geometric novelty F_{novelty} only, on the other hand, LLMs found textual prompts that yield cars with highly unconventional or fragmented shapes that have high novelty but are impractical (Fig. 5 (c)). By combining all three terms as a weighted sum in the objective function, LLM-to-Phy3D can effectively produce novel 3D artifacts that satisfy the physical and domain constraints (Fig. 5 (d)).

Fig. 6 reveals the highlighted visual features indicating the geometric novelty between a reference car and a car generated with LLM-to-Phy3D. In the top-row heatmap images in the figure, the geometric differences between the surfaces of a 3D-generated car and a reference car are highlighted, indicating visual cues of high geometric novelty. These differences correlate to the overlay meshes of the cars presented in the bottom row. Such geometric differences serve as an indicator of the region of novelty, allowing LLM-to-Phy3D to reward 3D artifacts in the objective function that not only lead to the discovery of novel artifacts but also satisfy physical and domain constraints. Investigation on the effectiveness of this geometric novelty measure under different camera projections reveals an LLM-to-Phy3D improvement with orthographic projection (*DPAR* $+3.38\%$ to $+52.21\%$). This indicates that the geometric distortions introduced by the projection strategy considerably affect the identification of physically conforming 3D artifacts. These proposed mechanisms enable LLM-to-Phy3D to find effective text prompts that yield novel, physically conforming designs (Fig. 7).

V. Conclusion

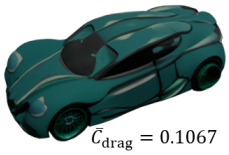
We introduce *LLM-to-Phy3D*, a novel physically-conforming text-to-3D online guidance method with LLMs. Our proposed

Prompt: "A car in the shape of a sleek dragon, with elongated, sweeping curves reminiscent of shimmering scales, illuminated LED eyes that emit a soft glow, wings that fold neatly against its body, and a tail that elegantly trails behind, embodying both speed and mythical elegance."

$\bar{C}_{\text{drag}} = 0.2027$

(a) Baseline

Prompt: "A car in the shape of a sleek, swirling illusion, adorned with whimsical currents that evoke the gentle flow of shimmering echoes, showcasing a matte jade hue that captures the essence of a surreal dusk, and headlights resembling softly flickering phantoms that shimmer in the twilight."



$\bar{C}_{\text{drag}} = 0.1067$

(b) LLM-to-Phy3D

Fig. 7. Examples of 3D cars generated with baseline and with LLM-to-Phy3D methods. Note that the lower the aerodynamic drag, the better the physical performance of the generated car.

method enables existing LLM-to-3D models to produce physically conforming 3D objects in the target domain on the fly. Central to LLM-to-Phy3D is an online black-box iterative refinement framework that adaptively steers the LLM to find textual prompts that yield physically plausible 3D artifacts with high geometric novelty. Systematic evaluations, supported by ablation studies of LLM-to-Phy3D in an aerodynamic vehicle design optimization scenario, reveal that our proposed method outperforms existing LLM-to-3D models in producing physically conforming target-domain 3D artifacts by more than 78%. Furthermore, the precise surface topology of 3D-generated objects, captured via physical light simulation with orthographic projection, enables effective measurement of geometric novelty relative to reference objects. These results underscore the potential general use of LLM-to-Phy3D in Physical AI for further scientific and engineering applications.

ACKNOWLEDGMENT

Melvin Wong gratefully acknowledges the financial support from Honda Research Institute Europe (HRI-EU). This research is also partly supported by the National Research Foundation, Singapore, and DSO National Laboratories under the AI Singapore Programme (AISG Award No.: AISG2-GC-2023-010, "Design Beyond What You Know: Material-Informed Differential Generative AI (MIDGAI) for Light-Weight High-Entropy Alloys and Multi-functional Composites (Stage 1b)", the College of Computing & Data Science (CCDS), Nanyang Technological University (NTU).

REFERENCES

[1] C. Dong, Y. Li, H. Gong, M. Chen, J. Li, Y. Shen, and M. Yang, "A survey of natural language generation," *ACM Comput. Surv.*, vol. 55, no. 8, Dec. 2022. [Online]. Available: https://doi.org/10.1145/3554727

[2] F. Bie, Y. Yang, Z. Zhou, A. Ghanem, M. Zhang, Z. Yao, X. Wu, C. Holmes, P. Golnari, D. A. Clifton, Y. He, D. Tao, and S. L. Song, "Renaissance: A survey into ai text-to-image generation in the era of large model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 3, pp. 2212–2231, 2025.

[3] H. Lee, M. Savva, and A. X. Chang, "Text-to-3d shape generation," in *Computer Graphics Forum*, vol. 43. Wiley Online Library, 2024, p. e15061.

[4] X. Li, Q. Zhang, D. Kang, W. Cheng, Y. Gao, J. Zhang, Z. Liang, J. Liao, Y. Cao, and Y. Shan, "Advances in 3d generation: A survey," *ArXiv*, vol. abs/2401.17807, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267334657

[5] M. Li, Y. Duan, J. Zhou, and J. Lu, "Diffusion-sdf: Text-to-shape via voxelized diffusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 12 642–12 651.

[6] H. Jun and A. Nichol, "Shap-e: Generating conditional 3d implicit functions," *arXiv preprint arXiv:2305.02463*, 2023.

[7] J. Xiang, Z. Lv, S. Xu, Y. Deng, R. Wang, B. Zhang, D. Chen, X. Tong, and J. Yang, "Structured 3d latents for scalable and versatile 3d generation," *arXiv preprint arXiv:2412.01506*, 2024.

[8] Z. Wang, J. Lorraine, Y. Wang, H. Su, J. Zhu, S. Fidler, and X. Zeng, "Llama-mesh: Unifying 3d mesh generation with language models," *arXiv preprint arXiv:2411.09595*, 2024.

[9] C. Banerjee, K. Nguyen, C. Fookes, and K. George, "Physics-informed computer vision: A review and perspectives," *ACM Computing Surveys*, vol. 57, no. 1, pp. 1–38, 2024.

[10] D. Liu, J. Zhang, A.-D. Dinh, E. Park, S. Zhang, A. Mian, M. Shah, and C. Xu, "Generative physical ai in vision: A survey," *arXiv preprint arXiv:2501.10928*, 2025.

[11] Q. S. Xu, J. Liu, M. Wong, C. Chen, and Y. S. Ong, "Looks great, functions better: Physics compliance text-to-3d shape generation," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2025.

[12] T. Rios, S. Menzel, and B. Sendhoff, "Large language and text-to-3d models for engineering design optimization," in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2023, pp. 1704–1711.

[13] M. Wong, T. Rios, S. Menzel, and Y. S. Ong, "Prompt evolutionary design optimization with generative shape and vision-language models," in *2024 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2024, pp. 1–8.

[14] L. Shimabucoro, S. Ruder, J. Kreutzer, M. Fadaee, and S. Hooker, "Llm see, llm do: Guiding data generation to target non-differentiable objectives," *arXiv preprint arXiv:2407.01490*, 2024.

[15] A. Nie, C.-A. Cheng, A. Kolobov, and A. Swaminathan, "Importance of directional feedback for llm-based optimizers," in *NeurIPS 2023 Foundation Models for Decision Making Workshop*, 2023.

[16] A. Jesson, N. Beltran Velez, Q. Chu, S. Karlekar, J. Kossen, Y. Gal, J. P. Cunningham, and D. Blei, "Estimating the hallucination rate of generative ai," *Advances in Neural Information Processing Systems*, vol. 37, pp. 31 154–31 201, 2024.

[17] J. Li, D. Li, S. Savarese, and S. Hoi, "Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," in *International conference on machine learning*. PMLR, 2023, pp. 19 730–19 742.

[18] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[19] S. G. Parker, J. Bigler, A. Dietrich, H. Friedrich, J. Hoberock, D. Luebke, D. McAllister, M. McGuire, K. Morley, A. Robison *et al.*, "Optix: a general purpose ray tracing engine," *Acm transactions on graphics (tog)*, vol. 29, no. 4, pp. 1–13, 2010.

[20] C. Rothwell, J. Mundy, and B. Hoffman, "Representing objects using topology," in *International Workshop on Object Representation in Computer Vision*. Springer, 1996, pp. 79–108.

[21] J. Montagnat, H. Delingette, and N. Ayache, "A review of deformable surfaces: topology, geometry and deformation," *Image and vision computing*, vol. 19, no. 14, pp. 1023–1040, 2001.

[22] K. Pulli and L. G. Shapiro, "Surface reconstruction and display from range and color data," *Graphical Models*, vol. 62, no. 3, pp. 165–201, 2000.

[23] P. Dutre, P. Bekaert, and K. Bala, *Advanced global illumination*. AK Peters/CRC Press, 2018.

[24] H. Lomax, T. H. Pulliam, D. W. Zingg, T. H. Pulliam, and D. W. Zingg, *Fundamentals of computational fluid dynamics*. Springer, 2001, vol. 246.

[25] M. H. Zawawi, A. Saleha, A. Salwa, N. Hassan, N. M. Zahari, M. Z. Ramli, and Z. C. Muda, "A review: Fundamentals of computational fluid dynamics (cfd)," in *AIP conference proceedings*, vol. 2030, no. 1. AIP Publishing, 2018.

[26] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby, "A tensorial approach to computational continuum mechanics using object-oriented techniques," *Computers in physics*, vol. 12, no. 6, pp. 620–631, 1998.