UDC 004.932.2:004.4’2 **doi: 10.32620/reks.2025.3.02**

Olesia BARKOVSKA, Liubov BUKHAROVA, Igor RUBAN,
Oleksii LIASHENKO, Andriy KOVALENKO, Vitalii SERDECHNYI

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

THE IMPACT OF THE NUMBER REPRESENTATION FORMAT IN PARAMETERIZED TEXT QUERIES ON THE ACCURACY OF 3D MODEL GENERATION

The subject matter *of the article is the analysis of the artificial intelligence models for generating 3D objects based on a text query with different formats for representing numerical parameters.* ***The goal of the work*** *is to evaluate and compare the effectiveness of the Hunyuan3D 2.0, Michelangelo and SDFusion models depending on the format of the text input queries using the following metrics: generation time, CLIP-Similarity and Chamfer Distance. The* ***tasks*** *to be solved are: 1) to conduct a systematic review of modern generative 3D models regarding their ability to process queries with numerical parameters presented in different formats (decimals, verbal representations, fractions); 2) to conduct a quantitative and qualitative evaluation of generative 3D models using the metrics of generation time, CLIP-Similarity and Chamfer Distance, taking into account the visual similarity to text descriptions and geometric similarity to Ground Truth objects; 3) to analyse the obtained results of generative text-to-3D models evaluation for parameterized 3D objects generation tasks for further implementation in user interface applications.* ***The methods used*** *in this work are machine learning methods, methods of vector representation of text and images, statistical methods of evaluating results, and a heuristic method of forming text queries.* ***The results*** *show that there is no universal generative 3D model that is capable of creating objects that fully correspond to a parameterised text query. The proposed methodology enables the formation of input text sequences containing such numerical representations, including decimals, verbal representations or fractions, for further analysis of the accuracy of object generation.* ***Conclusions****. Similar efficiency of training 3D generative models in the joint latent space of text features containing numerical parameters and using datasets with precisely defined geometric characteristics of objects was confirmed. The results show that the Hunuyan3D-2.0 model is suitable for further research and modification to adapt the used methods to create personalised 3D objects with given numerical parameters, such as a prosthetic cover.*

Keywords: *GenAI; similarity; precision; parameters; model; Chamfer Distance; CLIP-Similarity; text-to-3D.*

1. Introduction

The extension of leading deep learning methods to 3D spaces opens up a number of opportunities for both creativity and entertainment, as well as for business, namely design, prototyping, modelling and the use of generated models in the gaming industry and manufacturing. The characteristic feature of generative artificial neural networks (GenAI) is the ability to use models trained on large amounts of data and adapted to new use cases, ensuring their flexibility and scalability. The main task of 3D generative models is obvious – the creation of objects that meet specific requirements. For example, for medical and commercial applications, it is crucial that such models can utilise anthropometric measurements to create personalised 3D models. At the same time, other tasks require the generation of models with additional input parameters, such as material, lighting or physical properties of the desired object. Currently, leading solutions (Shap-E, LumaAI, Meshy, Sculpt3D, Dreamfusion) are mainly based on image or text query analysis, however, in order to improve the quality and accuracy of reconstructions, there are methods [1,2] in which additional input data, such as material or texture maps are incorporated. These approaches can be viewed as parameterized input queries, where each parameter constrains the generative process, but they also introduce the challenge of error estimation and balancing between constraints during model generation. Particularly, arises the need of accurately interpreting and mapping numerical inputs from text queries into meaningful 3D model parameters that more preserved with different text formulations. For instance, in industrial design, friction coefficients, tensile strength, or reflectance maps may be specified to ensure that the generated prototype not only has the correct appearance but also behaves realistically in simulation. Similarly, in digital content creation, lighting conditions, surface roughness, and spectral reflectivity can be included as parameters to ensure that the generated assets integrate seamlessly into existing rendering pipelines. Thus, direct integration of specialised parameters into the generation process is applied.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

1.1. Motivation

In today's world, when AI is becoming increasingly popular, the relevance of introducing machine learning methods into various areas of human activity is evident. More and more companies, approximately 15% of 2,000 organisations worldwide, according to the "AI in Action 2024" survey by IBM and The Harris Poll [3], are considered AI and technology leaders who improve aspects of their business, such as customer experience, IT operations and automation, virtual assistants, and cybersecurity. For example, two-thirds of leaders' reports [3] show that implementing AI has improved revenue growth rates by more than 25%, demonstrating the benefits of the latest technology in practice.

In particular, the development of generative AI, which uses powerful foundational models trained on large amounts of data, can be adapted to new use cases, ensuring the flexibility and scalability of the resulting product, which will significantly accelerate the adoption of AI. Thus, as of July 2024, according to 5,000 surveyed executives in 24 countries and 25 industries, 77% believe that generative AI is suitable for market implementation, compared to only 36% of those surveyed in 2023 [4].

1.2. State of the art

There are many works that highlight the latest achievements in the field of generating three-dimensional objects using artificial neural networks (ANNs). For example, one such work is the taxonomy of models into two classes, trained using 2D or 3D data [5]. In the study context, the authors review ANNs' learning approaches that model the data distribution of target 3D "templates" and support the use of training datasets for the synthesis of 3D representations and 2D images (the corresponding renders). In particular, the research focuses on existing methods used to represent three-dimensional shapes during the generation process and as the resulting object. These methods are evaluated by the efficiency of the time spent on display, memory efficiency of the methods themselves, the possibility of displaying objects in their present condition and compatibility with various NNs. These include voxel grids, point clouds, meshes and neural radiation fields (NERF). However, a newer representation method, Gaussian Splatting (GS)[6], which provides high-quality visualisation of neural fields for real-time tasks, is not considered.

In [7], a comprehensive analysis of existing generative NNs as of March 2024 is proposed and classifies 3D generative models into four general classes according to the data used during training. The 3DPT (Paired Text to 3D) class assumes the presence of a sample of paired text-3D data, 3DUT (Unpaired 3D Data) – models of this class use 3D representations that do not have defined labels, No3D (Text-to-3D without 3D data) – NNs that do not use three-dimensional objects as training data, and Hybrid3D – combine the other classes in some way. The use of diffusion or autoregressive priors over the latent space by the models is also considered.

The authors of the study of diffusion models in the context of 3D object creation [8] list all frequently used metrics in tasks related to 3D generation and classify them by proposing the following classes of indicators: distance, coverage, distribution, similarity, quality and errors. However, it should be noted that some metrics, in the context of specific models, are considered as metrics of adjacent categories in the works considered below. For example, Frechet Inception Distance (FID), which refers to the distribution metric, is indicated as a quality metric in the ShapeCrafter model [9].

Before GenAI, complex tasks requiring an adaptive approach were solved by traditional (discriminative) AI models [10] used in planning, forecasting, classification, architectural design and analysis, and in tasks such as regression and pattern recognition. Non-generative AI includes models not characterised by the creation of new content and focuses on classification, regression, recognition, forecasting, and other tasks, a general representation of which is shown in Figure 1.

Each of the above classes of models can be used in several specialised scenarios, for example, models for generating audio data are used in tasks such as voice synthesis, cloning, and reconstruction [11], image generation models for tasks such as reproducing new perspectives, generating unique images from text descriptions, etc.

Methods that use artificially generated data [12], such as images, to conduct simulations or further training of intelligent systems are widespread. Since graphical information can be considered any type of data represented in a visual form [13], generative models can be divided into classes according to the work with visual representations of information: two-dimensional and three-dimensional representations.

For the purposes of this study, it is essential to understand the concept of "Text-to-3D", which represents a class of generative AI models that, using methods of various modalities (such as learning from 3D objects, images, or mixed datasets) to create three-dimensional objects using the user's input text query (prompts) as a generation condition.

A comparative assessment of modern generative NNs for synthesising three-dimensional objects from text descriptions is given in Table 1.

In accordance with the results presented (recommended by the authors), we calculated the average similarity values estimated by the CLIP (Contrastive Language-Image Pre-training) model [14], which compares latent representations of renders with the corresponding

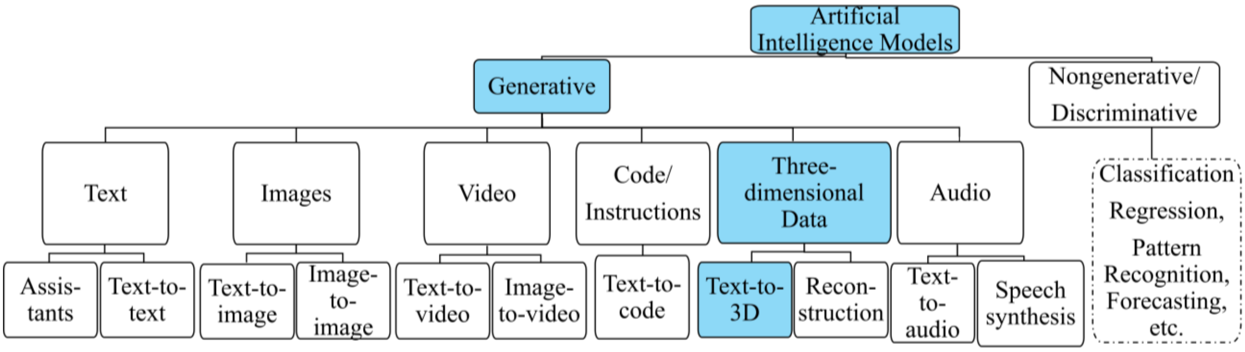


Fig. 1. General characteristics of artificial intelligence models by the types of resulting generation

textual description features. That was popularised in recent years [18] and is now commonly used for 3D models assessment in accordance with its textual description.

In the evaluation of "Text-to-3D" generative models, several metrics have become standard. Chamfer Distance (CD) and Earth Mover's Distance (EMD) are the most widely employed for measuring geometric fidelity of reconstructed point clouds or meshes. Light Field Distance (LFD) is also frequently applied, particularly where multi-view image-based comparison is more practical than direct geometry. Distribution-level measures such as Coverage (COV) and Minimum Matching Distance (MMD) are consistently used in generative modelling works to assess diversity and quality trade-offs. For perceptual evaluation of renderings, Frechet Inception Distance (FID) remains the most common choice, occasionally complemented by Frechet Video Distance (FVD) in studies using animated turntables of generated objects.

Other metrics include Kernel Inception Distance (KID) and Intersection over Union (IoU), which appear typically in ablation studies or voxel-based tasks; visual appearance similarity scores such as Visual Quality (VQ), Perceptual Quality (PQ) and Structure Completeness (SC), which are reported in more recent diffusion-based pipelines but not yet standardised. Text-alignment measures such as Prompt Fidelity (PF), which have been introduced in the context of CLIP-based evaluation of cross-modal consistency. Metrics drawn from 2D image or video generation, including Deep Image Structure and Texture Similarity (DISTS), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM), are also adopted to quantify rendering quality but are generally secondary to geometry.

While geometry-based measures such as Earth Mover's Distance (EMD) and Chamfer Distance (CD) provide rigorous numerical comparisons of 3D point distributions, they often fail to capture semantic fidelity. EMD is computationally expensive and sensitive to sampling, while CD may assign low error even when key structures are missing. Conversely, CLIP-S directly evaluates whether rendered views of a shape align with the input text description, capturing semantic and perceptual correctness. Among geometric metrics, CD is generally preferred over EMD in the following evaluation due to its lower computational cost and better scalability to large point clouds.

Taken together, CD and CLIP-S provide complementary guarantees: CD ensures numerical and structural compatibility, while CLIP-S ensures visual and textual accordance. This dual evaluation has therefore emerged as the most robust option for assessing "Text-to-3D" models dedicated to precise parameters integration, balancing geometric accuracy with semantic faithfulness to the prompt.

Thus, quantitative assessments of the generative capabilities of the analysed models, the scores considered were accuracy and similarity of three-dimensional objects.

The CLIP-S score was used as a similarity score, and the Chamfer Distance (CD) was used as an accuracy score in accordance with the numerical parameters. The CLIP-S metric is also used when comparing the results presented in Tables 1 and 2.

CLIP-S measures the average cosine similarity (normalised dot product) between the CLIP text embedding of the input description and the CLIP image embedding of the rendered views of the generated 3D shape:

$$\mathrm{CLIP-S}(c_i, e_i) = \frac{1}{N}\sum_{i=1}^{N}\frac{{c_i}^T e_i}{\|c_i\|\|e_i\|}, \qquad (1)$$

where c_i, – subregion image feature vector corresponding to the i-th word of the description sentence;

N – batch size;

T – the number of words;

$e \in \mathbb{R}^{D\times T}$ – matrix of text features, where D is the dimension of the word vector.

Physically interpretable as the normalised projection of one vector onto another. Thus, a higher value means stronger alignment of semantics across modalities.

However, CLIP-S is not consistent when comparing the similarity of models with textures and exclusively ge-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

ometry, which can be seen in the example of the Hunyuan3D 2.0 model, which received a score of 0.3751 when checking an object without texture, which is by 0.0425 less than the result of a "complete" object.

Also evident are higher scores when comparing renders to input images, indicating the tendency of the CLIP-S metric to focus primarily on general visual features, ignoring significant differences in the representation of 3D structures.

This being said, to analyse the geometry of the generated objects, the chamfer distance (CD) metric was used, which is a known metric for quantifying the difference between two point clouds [35]:

$$\mathrm{CD}(x_i, y_i) = \frac{1}{|x_i|}\sum_j \min_k d(x_{ij}, y_{ik}) + \frac{1}{|y_i|}\sum_k \min_j d(x_{ij}, y_{ik}) \quad (2)$$

where x_i and y_i , – vectors of point clouds of the objects;

$d(\cdot,\cdot)$ – distance metric, in this study, Euclidean distance was used.

From a geometric perspective, CD is the average minimal distance between two point clouds/surfaces. According to this, lower values indicate greater similarity.

The generation process from the text description of the LGM, DreamGaussian and MicroDreamer models, is based on the image synthesised by the pre-trained Text-to-Image (T2I) model. Hence, the corresponding text prompts using GPT-4o LLM were generated to deliver the description of objects to the CLIP assessment. This need arises due to the possible contradictory results [15] in the case of direct comparison of image and text similarity scores with the values of the image-render scores.

To find out existing connections between the presented models, hence the rationality of their direct comparison, in the case of conducting research on the specified categories of objects, a Venn diagram was constructed (Fig. 2).

Figure 2 shows stable relationships between the classes of generated objects and the models, covering a wide range of resulting forms. Accordingly, the high degree of generalizability of the studied models allows us to compare models based on queries from different semantic categories.

It is worth noting that obtaining direct renders from three-dimensional objects was carried out in the Hunyuan3D 2.0 and CLAY models, using the Python kiui package, in other cases the studied images were obtained by sampling from officially presented animation videos, with subsequent removal of the background and reduction of frames to a single resolution (96 dpi) and size (512×512 pixels).

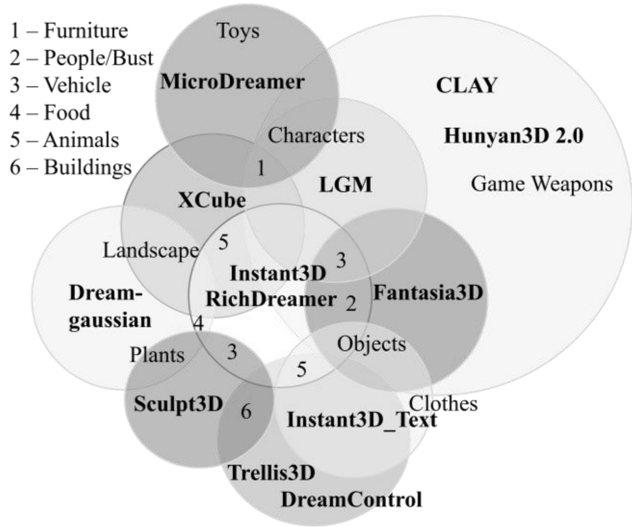


Fig. 2. Venn diagram of connections between generative NMs for synthesising three-dimensional objects from textual description

Michelangelo, Craftsman, and SDFusion models were compared separately for generating object geometry and are presented in Table 2, since these models do not generate textures.

The comparative analysis was carried out based on the averaged CLIP-Similarity (CLIP-S) score, as proposed by the authors of the MicroDreamer model [19]: the CLIP ViT-bigG/14 – LAION-2B model, trained on 5.85 billion filtered image-text pairs, was used as a visual similarity score. CLIP-S is calculated for eight object views, for 3 different object categories, evenly distributed in azimuth angles, starting from 0. In total 24 samples were used for the experiment. The Trellis and Sculpt3D models are different, where the views were collected unevenly. The objects for the study were selected to test the ability to generate generalised forms of common query categories (animals, buildings, furniture, figures, busts, etc.).

An important aspect is that only the authors of the Michelangelo model include text queries containing numerical values in their test prompt examples. In contrast, other models do not check this factor (the correspondence of generations to numerical dimensions).

The following models also performed well: DreamGaussian, LGM, and MicroDreamer, but were excluded from further study due to their pre-generated image training methodology, which complicates the process of integrating precise digital parameters. The CLAY and Instant3D models do not have an open-source implementation, while RichDreamer and Fantasia3D require significant computing power and time to generate a single object.

Table 1

Comparative analysis of generative NMs for the synthesis of three-dimensional objects from text description

Model	Mean CLIP-S	Generation Time, s	Classes of generation
LGM [16]	0.3933	3	“Characters”, “Furniture”,” Vehicle”
DreamGaussian [17, 18]	0.4078	120	“Food”, “Landscape”, “Plants”
MicroDreamer [19]	0.4220	29	“Characters”, “Toys”,” Furniture”
Trellis3D [20]	0.3942	10	“Buildings”, “Animals”, “Objects”
Fantasia3D [21, 18]	0.4249	1860	“Vehicle”, “Objects”, “People/Bust”
RichDreamer [22]	0.4625	5400	“Vehicle”, “Food”, “Animals”
Instant3D [23]	0.4794	20	“Food”, “Animals”, “Vehicle”
Instant3D: Instant Text-to-3D Generation [24]	0.4049	0,025 – 1	“Animals”, “Clothes”, “People/Bust”
Sculpt3D [25]	0.3907	–	“Vehicle” “Plant”, “Building”
XCube [26]	0.3979	30	“Landscape”, “Animals”, “Furniture”
DreamControl [27]	0.4021	1800	“Animals”, “Buildings”, “Objects”
CLAY [28, 29]	0.4341	–	“Objects”, “Game weapons”, “Furniture”
Hunyuan3D 2.0 [30]	0.4176	30	“Objects”, “Game weapons”, “Furniture”

Table 2

Comparative geometry analysis

Model	Mean CLIP-S	Generation Time, s	Classes of generation
Michelangelo [31]	0.4101	3 – 10	“Animals”, “Toys”, “Furniture”
Craftsman [32]	0.3745	5 – 30	“People/Bust”, “Characters”, “Buildings”
Trellis3D [20]	0.3296	10	“Buildings”, “Animals”, “Objects”
CLAY [30]	0.3646	–	“Objects”, “Game weapons”, “Furniture”
Hunyuan3D 2.0 [30]	0.3751	30	“Objects”, “Game weapons”, “Furniture”
SDFusion [33]	0.3313	–	“Furniture”, “Primitive Shapes”

The Instant3D model (Instant Text-to-3D Generation) aims to generate externally accurate three-dimensional models, but with fuzzy and poorly detailed geometry. This model also uses the CLIP text encoder, which may give it an advantage in the similarity analysis called homonymous.

Michelangelo and Craftsman were considered as baseline models for comparison and analysis in the paper in which the Hunyuan3D 2.0 model was first proposed. According to the results presented in [30], Hunyuan3D 2.0 demonstrates the highest quality of generated 3D meshes when using an image as a generation condition. However, proving that it maintains the functionality of the created objects and compliance with numerical parameters when using text queries requires additional research.

Thus, for further practical analysis, open source models were selected – Michelangelo, Hunyuan3D 2.0 and SDFusion; among which Michelangelo and Hunyuan3D 2.0 are distinguished by higher accuracy rates in relation to generation time according to the CLIP-S score, and SDFusion was chosen as a representative trained exclusively on the ShapeNet dataset, which contains real-size annotations.

1.3. Aim and objectives

The aim of the paper is to study the efficiency of generative 3D AI models depending on the format of the text query, namely, the form of representation of numerical parameters for generating 3D objects and further implementation in automated parameterised generative systems.

To achieve this aim, the following objectives have been identified:

– a systematic review of modern 3D generative models (in particular, Hunyuan3D 2.0, Michelangelo, SDFusion) regarding their ability to process queries with

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

numerical parameters presented in different formats (decimals, verbal representations, fractions);

– quantitative and qualitative evaluation of generative 3D models by the metrics of generation time, CLIP-Similarity and Chamfer Distance, considering visual similarity to textual descriptions and geometric similarity to Ground Truth objects;

– analysis of the obtained results of the evaluation of text-to-3D generative models for the tasks of parameterised generation of 3D objects for further implementation in user interface applications.

The prospect of using the obtained research results is to implement them in the protection subsystem of prosthetic limbs by automated manufacturing of prosthetic covers for personalised generation of three-dimensional objects according to individual parameters.

The paper is structured as follows. Section 1 provides a review of recent GenAI models and conducts an evaluation of them by visual object accuracy, resulting in a set of top NNs for further analysis. Section 2 (Materials and methods of research) discusses the ways of input prompt formation strategy and emphasises advances and limitations of metrics CLIP-S and Chamfer Distance in the context of objects' precision evaluation. Section 3 (Results and Discussion) provides results of a comprehensive analysis identifying a NN model potentially capable of generating personalised objects – Hunyuan3D 2.0 via performance analysis of the Hunyuan3D 2.0, Michelangelo and SDFusion models depending on the format of the text input queries. Section 4 (Conclusions) presents the implications of the findings, including their applications in real-world scenarios such as the prosthetic limb protection subsystem, i.e. prosthetic cover creation workflow.

2. Materials and methods of research

As mentioned above, most text-to-3D models today implicitly check the results for the correspondence of generated objects to numerical parameters and use generalised queries focused on the model's appearance or understanding of latent textual features when generating from textual prompts of different lengths [34]. Thus, for the evaluation, the formation of text queries was designed to check both the geometry of the object and spatial relations in different representation format of the numerical parameters of one metric unit (meters), namely in the form of:

– decimals and whole numbers;

– fractions and whole numbers in verbal form;

– completely verbal representations of numbers.

Also, the described input text sequences have different complexity, depending on the length of the prompt sentences (queries): from 23 to 43 words, of which from 1 to 7 parameters, respectively.

The experiment is conducted for four objects of the "furniture" class, namely: "chairs", "tables" and "bookcases". Such coverage allows us to investigate the intra-class variability of shapes, sizes and structural features characteristic of furniture, which, in turn, is a necessary condition for the development of a neural network model capable of generating personalized objects within the same class, with dependence of the application (for example, prosthetic covers). Taking into consideration the need to output an object with a clear structure, the parameters in the queries are presented for individual components in brackets (Table 3).

Point cloud vectors are obtained from the vertices and surfaces of the objects under study.

Thus, during further research, the resulting objects are checked for compliance with the specified dimensions and text description.

The hypothesis of the experiment is that existing 3D generative models do not equally perceive the representation of numerical values in text queries.

As part of the experiment, a qualitative assessment is carried out on two-dimensional representations of the geometry of objects that are further used during the quantitative score CLIP-S testing (Tables 4, 5). Hunyuan3D-2.0, SDFusion and Michelangelo were chosen as the models because the analysis results (Tables 1, 2) showed that these models are distinguished among others by higher accuracy rates in relation to generation time and are potentially capable of rational processing and perception of numerical parameters based on the training data used.

To eliminate potential differences in the process of rendering, we used normal mapping, which provides a visualisation of the orientation of object surfaces.

Testing to detect the coherence of formed geometry was completed using CD by contrasting selected and modified, following the given numerical parameters, "true" 3D objects [36 - 38] with the generated forms.

Using the Blender software, "true" objects obtained from open access were edited and thus brought to Ground Truth (GT) of the current study context.

As can be seen from Figure 3, the 3D objects used in the quantitative comparison correspond to the description of the generated objects to tenths of the chosen unit system – meters (see Table 3). Thus, deviations of 0.01 meters are allowed.

In turn, due to the metric consistency of the objects, these 3D models are suitable for correctly calculating the distance between the generated and reference (GT) shapes using the CD metric, which is sensitive to spatial deviations at the level of surface points.

Table 3

The examples of parametrized requests with different formats of numerical parameters (decimals, verbal representations, fractions)

<table>
<tr><th>Request for testing selected models (the prompt)</th><th>Peculiarities of the text input queries</th></tr>
<tr><td>A slat back wooden chair of 0,9 meters with 4 legs and a seat (height: 0,45m, width: 0,4m), It has 5 vertical backrest slats spaced at 0,008m</td><td rowspan="3">The query contains 27/30/40 words, with the following parameters: decimals, fractions, and numbers as decimals in purely verbal form;
- 2 parameters as whole numbers and 4 in the form of decimal fractions;
- 2 parameters as whole numbers in verbal representation and 4 in the form of ordinary fractions.
- 6 parameters in verbal representation.
Understanding of basic spatial directions and mixed representation of units of measurement is inspected.</td></tr>
<tr><td>A slat back wooden chair of 9/10 meters with four legs and a seat (height: 9/2m, width: 2/5m), It has five vertical backrest slats spaced at 1/125 of a meter</td></tr>
<tr><td>A slat back wooden chair of zero point nine meters with four legs and a seat (height: zero point fourty-five meters, width: zero point four meters). It has five vertical backrest slats spaced at zero point zero zero eight meters</td></tr>
<tr><td>A round table with a tabletop (diameter: 0,8m), supported by 3 slender legs (height: 0,4m), The legs connect to a circular gold base ring</td><td rowspan="3">The query contains 24/24/30 words, with the following parameters: decimals, fractions, and numbers as decimals in purely verbal form;
- 1 parameter as whole numbers and 2 in the form of decimals;
- 1 parameter in whole numbers as verbal representation and 2 in the form of fractions.
- 3 numerical parameters in verbal form.
Inspected understanding of non-typical spatial relationships and generalised representation of units of measurement.</td></tr>
<tr><td>A round table with a tabletop (diameter: 4/5m), supported by three slender legs (height: 2/5m). The legs connect to a circular gold base ring</td></tr>
<tr><td>A round table with a tabletop (diameter: zero point eight meters), supported by three slender legs (height: zero point four meters). The legs connect to a circular gold base ring</td></tr>
<tr><td>A black rectangular bookcase (1.8m high, 0.9m wide, 0.3m deep). Include 6 evenly spaced shelves (0.28m apart, 0.02m thick), 1 wide drawer at the bottom</td><td rowspan="3">The query contains 24/24/43 words, with the following parameters: decimals, fractions, and numbers with decimals in purely verbal form;
- 2 parameters as whole numbers and 5 in the form of decimal fractions;
- 2 parameters as whole numbers in verbal representation and 5 in the form of ordinary fractions.
- 7 numerical parameters in verbal form.
The understanding of spatial relationships, relative adjectives and a generalised representation of units of measurement are inspected.</td></tr>
<tr><td>A black rectangular bookcase (9/5m high, 9/10m wide, 3/10m deep). Include six evenly spaced shelves (7/25m apart, 1/50m thick), one wide drawer at bottom</td></tr>
<tr><td>A black rectangular bookcase (one point eight meters high, zero point nine meters wide, zero point three meters deep), includes six evenly spaced shelves (zero point two eight meters apart, zero point zero two meters thick), and one wide drawer at the bottom</td></tr>
</table>

A general qualitative comparison of the original and generated objects is shown in Figure 4.

Considering the non-invariance of the Chamfer Distance metric to scale and to preserve the generated parametric relations, all objects studied during the experiment were normalised by aligning them in space (bringing them to a single scale).

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

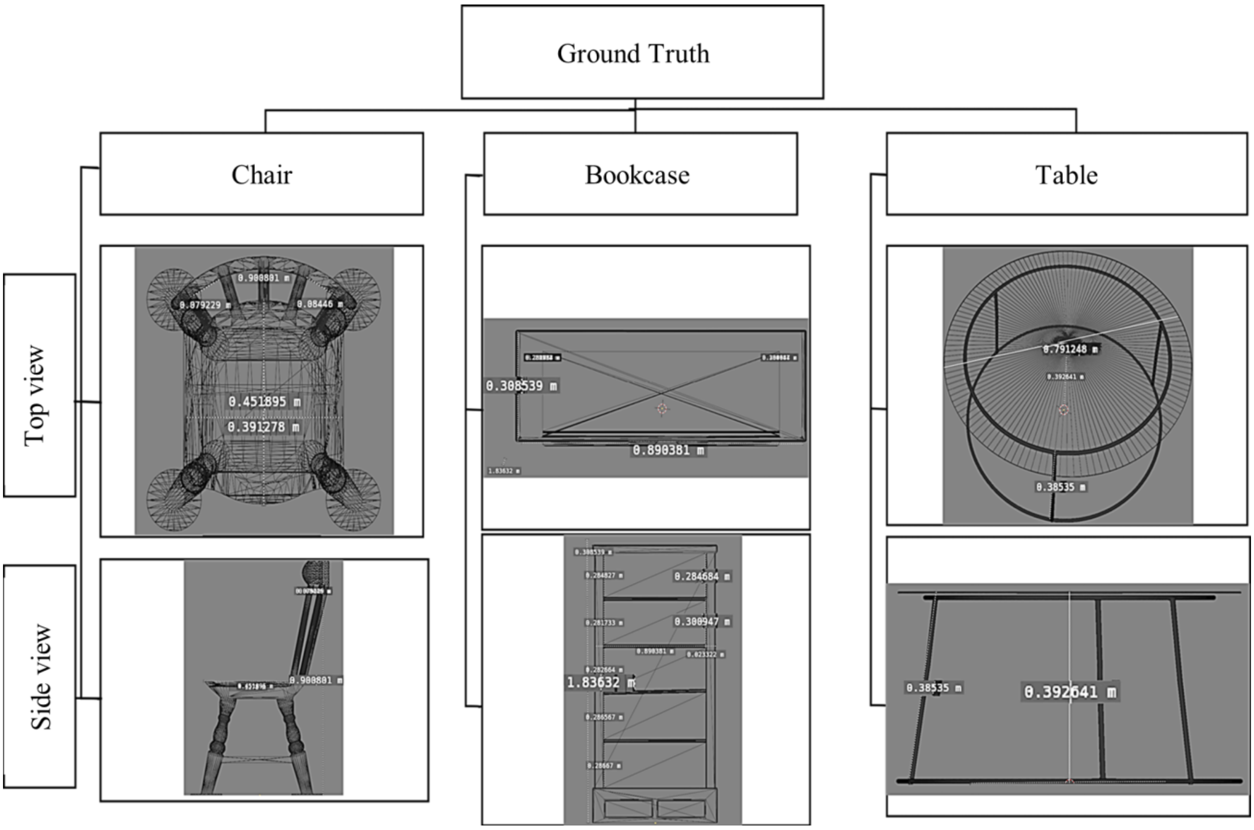
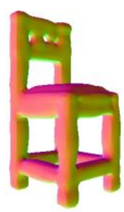
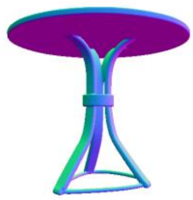
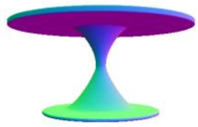
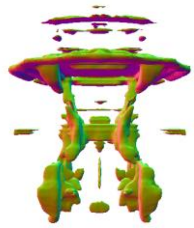
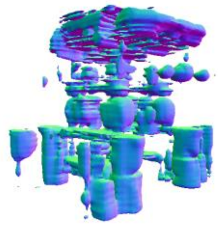
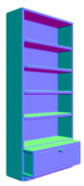
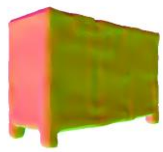
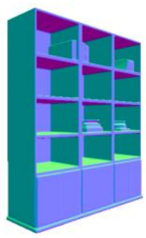
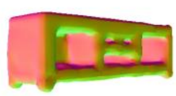


Fig. 3. Diagram of the angles corresponding to the Ground Truth objects

Table 4

Qualitative comparison between the objects

№	The prompt	Hunyuan3D 2.0	SDFusion	Michelangelo
Testing the models based on the chair prompt request				
1.1	A slat back wooden chair of 0.9 meters with 4 legs and a seat (height: 0.45m, width: 0.4m). It has 5 vertical backrest slats spaced at 0.008m.			
1.2	A slat back wooden chair of 9/10 meters with four legs and a seat (height: 9/2m, width: 2/5m). It has five vertical backrest slats spaced at 1/125 of a meter.			
1.3	A slat back wooden chair of zero point nine meters with four legs and a seat (height: zero point fourty-five meters, width: zero point four meters). It has five vertical backrest slats spaced at zero point zero zero eight meters.			

№	The prompt	Hunyuan3D 2.0	SDFusion	Michelangelo
	Testing the models based on the table prompt request			
2.1	A round table with a tabletop (diameter: 0.8m), supported by 3 slender legs (height: 0.4m). The legs connect to a circular gold base ring.			
2.2	A round table with a tabletop (diameter: 4/5m), supported by three slender legs (height: 2/5m). The legs connect to a circular gold base ring.			
2.3	A round table with a tabletop (diameter: zero point eight meters), supported by three slender legs (height: zero point four meters). The legs connect to a circular gold base ring.			
	Testing the models based on the bookcase prompt request			
3.1	A black rectangular bookcase (1.8m high, 0.9m wide, 0.3m deep). Include 6 evenly spaced shelves (0.28m apart, 0.02m thick), 1 wide drawer at the bottom			
3.2	A black rectangular bookcase (9/5m high, 9/10m wide, 3/10m deep) Include six evenly spaced shelves (7/25m apart, 1/50m thick), one wide drawer at bottom.			
3.3	A black rectangular bookcase (one point eight meters high, zero point nine meters wide, zero point three meters deep) includes six evenly spaced shelves (zero point two eight meters apart, zero point zero two meters thick) and one wide drawer at the bottom.			
	Mean generation time, s	90	65.5	19.5
	Equipment	NVIDIA L4 Tensor Core GPU	NVIDIA T4 Tensor Core GPU	NVIDIA L4 Tensor Core GPU

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Fig. 4. Representation of GT objects in accordance with the generated on

3. Results and Discussion

The results of the quantitative average, over the six processed examples of assessment of 2D representations in accordance with the text queries, as well as the chamfer distances, are given in Table 5.

As mentioned above, normalised scores give a higher index of similarity between objects.

From the qualitative results, it can be seen that most models are able to generate geometry under typical conditions (for example, a chair with four legs), while customized objects, such as a modern table, lead to results that are not identical to the input prompts or even generate disjointed, abstract geometry, as in the case of the Michelangelo model, which can significantly affect the further use of such models in printing tasks, where one of the requirements is the watertightness of the models.

Also noticeable are errors in generating spatial relationships – only the Hunyan3D-2.0 model could handle the task of generating a bookcase with a single wide bottom drawer, while deforming its appearance.

The quantitative evaluation results show that the Michelangelo model, even with visually opened geometry, still has a higher score than the SDFusion model.

This behaviour may be due to the specifics of testing using the CD. Since the distance is calculated based on all vertices of the generated shapes, differences in point density between models can lead to distorted results. In particular, 3D models that occupy a larger area or contain more geometric details have a higher number of points, which can result in higher metric values even if the geometric accuracy is visually lower. In accordance with this observation, additional testing was conducted, the results of which are presented in Table 6.

The points for testing were selected from the surfaces of the objects, with the maximum number of vertices in the object generated by the Michelangelo model: 104909, Hunyuan3D 2.0: 340269 and SDFusion: 19122.

The results are presented in Table 6, demonstrating that the number of sample points taken from the object surface does not have a significant impact on the value of the Chamfer Distance metric, which indicates its stability in the context of the surface discretisation sampling density.

The results (see Table 6) show that as the number of surface sampling points (vertices) increases, the chamfer distance values increase, so for further experiments, testing on 19122 surface points was selected as giving better results. Thus, the Hunyuan3D 2.0 model was chosen for further experiments as the one that stands out among others for the most accurately generated geometry and appearance of objects.

Table 5

Quantitative evaluation of the studied models using Chamfer Distance and CLIP-S metrics

The models	CLIP-S↑	CD (normalised)↓	CD↓
Hunyuan3D 2.0	**0.3639**	**0.1289**	**0.4044**
SDFusion	0.3089	0.2170	0.4964
Michelangelo	0.3182	0.1992	0.5029

Table 6

Quantitative assessment of the results of testing the studied models using the Chamfer Distance metric, depending on the number of sampling points

The models	CD (normalised)↓			
	Sample number of points: 19122	Sample number of points: 9561	Sample number of points: 4780	Sample number of points: all the vertices of the object
Hunyuan3D 2.0	0.6985	0.7020	0.7089	0.1289
SDFusion	0.1346	0.1351	0.1366	0.2168
Michelangelo	0.1214	0.1235	0.1246	0.1992

For a complete evaluation of the models under study, we present a comparative diagram of the generation time of 3D objects by the Hunyuan3D 2.0, SDFusion, and Michelangelo models based on the data given in Table 4 (Fig. 5).

It is evident that the Hunyuan3D 2.0 model, at the same time, has the highest indicators of the studied metrics and the highest timing values. The Michelangelo model demonstrates the shortest generation time among the considered methods, while maintaining a higher level of similarity than the SDFusion model by the CLIP-Similarity metric and a lower CD score.

However, when examining Hunyuan3D 2.0 for the accuracy of matching numerical parameters, you can see from Figure 6 that the model generates only approximately the same numerical relationships regardless of the parameters format set.

Figure 6b shows the results for the queries with decimals, words (verbal representation), and fractions, respectively, from left to right. When analysing the detailed query (Table 7), we see that only the first model (prompt with decimals) has six shelves and a wide drawer, as specified in the input query, while the height of the cabinet deviates from the query request by 0.75 meters. Thus, the smallest errors are observed in the measurements of shelf thickness, and the results with fractions included give the highest error of 0.89 meters.

The input numerical parameters and the format of the text query affect the accuracy of the metrics by which the generative model was evaluated. As such evident the need of accurately interpreting and mapping numerical inputs from text queries into meaningful 3D model parameters that more preserved with measurements presented in integer format decimals and fractional values given in verbal forms as the results of the experiment revealed. The most suitable for further improvement for specialised applications is the generative model Hunyuan3D 2.0.

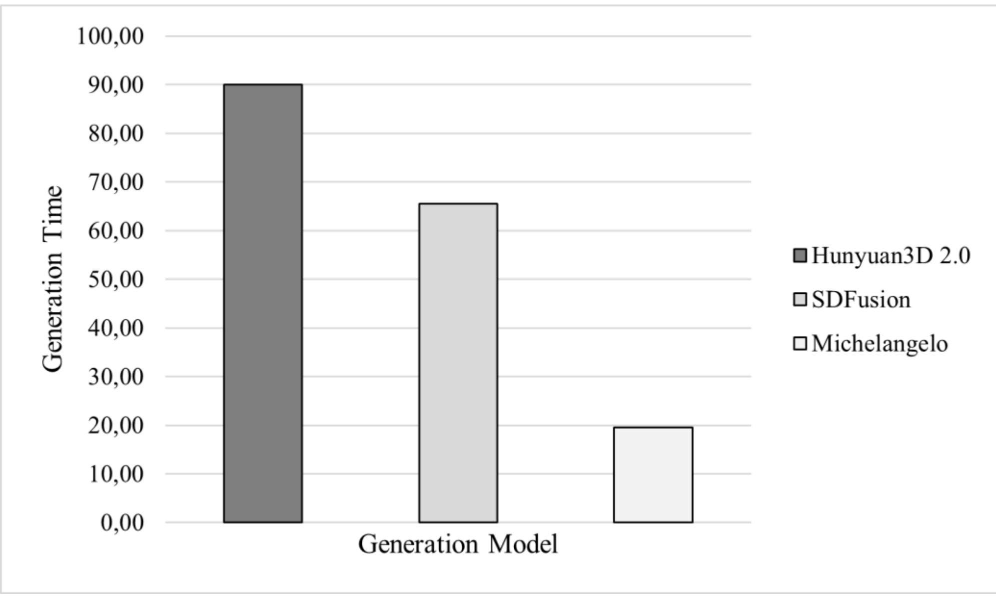
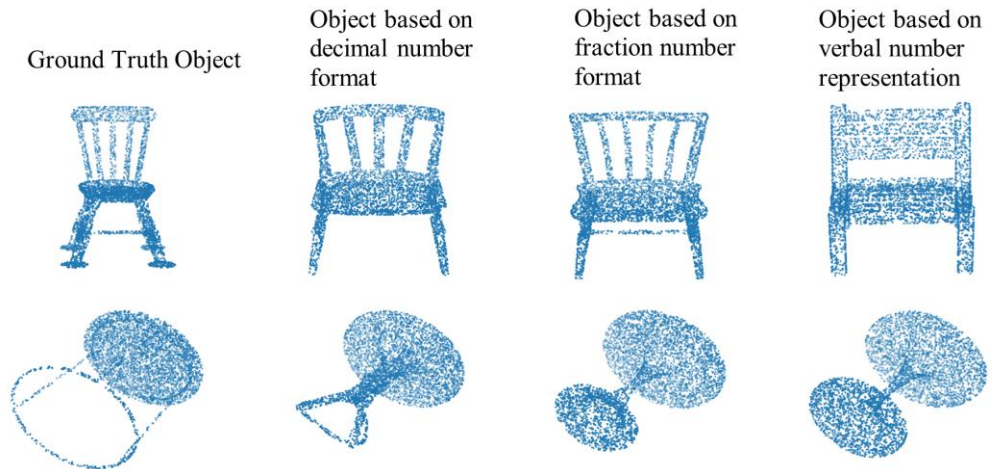
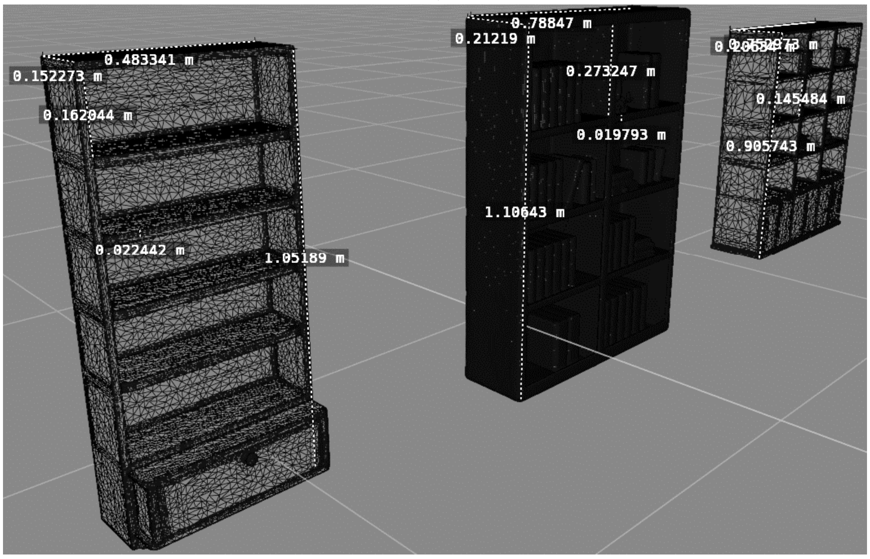


Fig. 5. Correlation between the generation time of 3D objects and the selected generative model

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)



a)



b)

Fig. 6. Qualitative comparison of normalised objects generated by the Hunyuan3D model of four types of queries: a) display by sample points, b) detailed comparison in Blender software

Table 7

Errors in numerical parameters of the generated Hunyuan3D 2.0 object: "Bookcase"

Set values		Error		
		Using decimal parameters (model 1)	Using verbal parameters (model 2)	Using fraction parameters (model 3)
Heigth	1.8	0.75	0.68	0.89
Width	0.9	0.42	0.12	0.15
Depth	0.3	0.15	0.088	0.094
Distance between shelves	0.28	0.12	0.01	0.13
Thickness	0.02	0	0	0.002

4. Conclusions

The work investigates the efficiency of generative 3D models depending on the format of the input text query, namely, the form of representation of numerical parameters for the generation of 3D objects. The efficiency was evaluated by the following metrics: object generation time, CLIP-Similarity (visual similarity), and Chamfer Distance (geometric similarity to the Ground Truth object).

As a result of the research and experiments conducted, potentially suitable models for modernisation to improve the compositionality and consistency of the object in accordance with the text query (prompts) – Hunyuan3D 2.0, SDFusion, and Michelangelo, were analysed.

Calculating Chamfer Distance when specifying a specific number of sampling points from the surface (19122, 9561, 4780) showed that the internal consistency of the results between the three models under study was maintained, i.e. the generated objects do not have a significant number of internal artifacts that could potentially cause disruption of the object functionality and unnecessary expenses during further processing of the 3D models.

The analysis of generative models on the NVIDIA L4 (T4) Tensor Core GPU showed that Michelangelo demonstrates the shortest generation time among the considered methods, 78.3% faster than Hunyuan3D 2.0 and 70.2% faster than SDFusion, while maintaining a higher level of similarity by the CLIP-Similarity metric and a lower CD score. This result demonstrates its effectiveness in tasks where the speed of generation is critical, with a moderate quality of matching the text description. The stability of Chamfer Distance values confirms the usefulness of using precise text annotations containing numeric values during GenAI model training. At the same time, the CD (19122 sampling points) results of the SDFusion model are 0.013144 inferior to the Michelangelo model, which is an insignificant difference compared to the gap of the Hunyuan3D 2.0 model (42.48% less than the SDFusion). This may indicate similar learning efficiency as for shared latent space, with numerical parameters in text descriptions (Michelangelo) and, at the same time, in the use of datasets with clearly defined geometric characteristics of objects (SDFusion).

Thus, the conducted research is worth applying to use as a knowledge base for problem-solving tasks related to generating a specific object that must meet precise requirements in terms of parameters. The Hunuyan3D-2.0 model and its method for creating the geometry of three-dimensional shapes using the Transformer architecture are suitable for further research and modification to be adapted to create personalised three-dimensional objects according to given numerical parameters, such as prosthetic covers.

The study's results are going to be implemented in the protection subsystem of prosthetic limbs by the automated manufacturing of prosthetic covers for personalised generation of three-dimensional objects according to individual user parameters. A wide range of generatable three-dimensional object classes, as well as high accuracy rates according to the CD and CLIP-S metrics, prove the feasibility and validity of using the Hunuyan3D-2.0 model for implementation in the prosthetic limb protection subsystem, but do not limit it to this task.

Further development of the subsystem will be carried out within the framework of the research project "Integrated technological solutions for accelerating the physical and socio-psychological rehabilitation of prosthetic patients" (state registration number: 0125U001654).

Contribution of authors: formulation of the problem – **Olesia Barkovska**. **Liubov Bukharova**, **Igor Ruban**; hardware platform investigation and development tools selection – **Liubov Bukharova, Igor Ruban**.

Methodology of the experiments – **Olesia Barkovska**; **Andriy Kovalenko,** data collection and preprocessing – **Liubov Bukharova**, **Vitalii Serdechnyi**; analysis and processing of the obtained results, visualisation – **Olesia Barkovska**, **Igor Ruban**, **Oleksii Liashenko**; finalisation of the draft article version – **Olesia Barkovska**; **Andriy Kovalenko,** corrections and postediting – **Oleksii Liashenko**.

Conflict of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

This study was conducted with financial support from the research project "Integrated technological solutions for accelerating the physical and socio-psychological rehabilitation of prosthetic patients" (state registration number: 0125U001654).

Data Availability

Data will be made available upon reasonable request.

Use of Artificial Intelligence

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, as described in the research methodology section.

Acknowledgments

The work was supported by the state-funded research project of Kharkiv National University of Radio Electronics entitled "Integrated technological solutions for accelerating physical and socio-psychological rehabilitation of prosthetic patients" (state registration number 0125U001654).

All the authors have read and agreed to the published version of this manuscript.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

References

1. Barkovska, O., Oliinyk, D., Sorokin, A., Zabroda, I., & Sedlaček, P. A system for monitoring the progress of rehabilitation of patients with musculoskeletal disorder. *Advanced Information Systems*, 2024, vol. 8, no. 3, pp. 13–24. DOI: 10.20998/2522-9052.2024.3.02.

2. Bukharova, L. D., & Barkovska, O. Y. 3D-generatziya z vykorystannyam opysovykh i chyslovykh zapytiv [3D-generation using descriptive and numerical queries]. *Suchasni napryamy rozvytku informatziino-communicatziinykh tekhnologii ta zasobiv upravlinnya – Current Directions of Development of Information and Communication Technologies and Control Tools*, 2025, vol. 3, p. 26. DOI: 10.32620/ict.25.t3. (In Ukrainian).

3. *AI in action 2024 report | IBM. IBM - United States*, 2024. Available at: https://www.ibm.com/think/reports/ai-in-action (accessed 31.05.2025).

4. Goehring, B., Goyal, M., Gunnar, R., Marshall, A., & Soffer, A. *The Ingenuity of Generative AI at Scale*, 2024. Available at: https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/scale-generative-ai. (accessed 31.05.2025).

5. Shi, Z., Peng, S., Xu, Y., Geiger, A., Liao, Y., & Shen, Y. Deep Generative Models on 3D Representations: A Survey, *arXiv*, 2023, DOI: 10.48550/arXiv.2210.15663.

6. Kerbl, B., Kopanas, G., Leimkuehler, T., & Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 2023, vol. 42, no. 4, pp. 1–14. DOI: 10.1145/3592433.

7. Lee, H., Savva, M., & Chang, A. X. 7.Text-to-3D Shape Generation. *Computer Graphics Forum*, 2024, vol. 43, no. 2. DOI: 10.1111/cgf.15061.

8. Wang, Z., Li, D., & Jiang, R. Diffusion Models in 3D Vision: A Survey, *ArXiv*, 2024. DOI: 10.48550/arXiv.2410.04738.

9. Fu, R., Zhan, X., Chen, Y., Ritchie, D., & Sridhar, S. ShapeCrafter: A Recursive Text-Conditioned 3D Shape Generation Model. In: S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh, eds. *Advances in Neural Information Processing Systems. Curran Associates, Inc*., 2022, pp. 8882–8895. DOI: 10.5555/3600270.3600916.

10. Rashidi, H. H., Pantanowitz, J., Hanna, M. G., Tafti, A. P., Sanghani, P., Buchinsky, A., Fennell, B., Deebajah, M., Wheeler, S., Pearce, T., Abukhiran, I., Robertson, S., Palmer, O., Gur, M., Tran, N. K., & Pantanowitz, L. Introduction to Artificial Intelligence and Machine Learning in Pathology and Medicine: Generative and Nongenerative Artificial Intelligence Basics. *Modern Pathology*, 2025, vol. 38, no. 4, article no. 100688. DOI: 10.1016/j.modpat.2024.100688.

11. Kamath, P., Morreale, F., Bagaskara, P. L., Wei, Y., & Nanayakkara, S. Sound Designer-Generative AI Interactions: Towards Designing Creative Support Tools for Professional Sound Designers. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery*, 2024. DOI: 10.1145/3613904.3642040.

12. Nikolenko, S. I. Synthetic Data for Deep Learning. *Springer International Publishing*, 2021. DOI: 10.1007/978-3-030-75178-4.

13. Strothotte, T., & Schlechtweg, S. 7 – Geometric Models and Their Exploitation in NPR. In: T. Strothotte and S. Schlechtweg, eds. *Non-Photorealistic Computer Graphics. San Francisco: Morgan Kaufmann*, 2002, pp. 203–245. DOI: 10.1016/B978-155860787-3/50008-6.

14. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. Learning Transferable Visual Models From Natural Language Supervision, *ArXiv*, 2021. DOI: 10.48550/arXiv.2103.00020 (accessed 31.05.2025).

15. Huang, J.-H., Zhu, H., Shen, Y., Rudinac, S., & Kanoulas, E. Image2Text2Image: A Novel Framework for Label-Free Evaluation of Image-to-Text Generation with Text-to-Image Diffusion Models. *MultiMedia Modeling: 31st International Conference on Multimedia Modeling*, 2025, Nara, Japan, January 8–10, pp. 413–427. DIO: 10.1007/978-981-96-2071-5_30.

16. Tang, J., Chen, Z., Chen, X., Wang, T., Zeng, G., & Liu, Z. LGM: Large Multi-View Gaussian Model for High-Resolution 3D Content Creation. *ArXiv,* 2024. Available from: https://arxiv.org/abs/2402.05054 (accessed 31.05.2025).

17. Tang, J., Ren, J., Zhou, H., Liu, Z., & Zeng, G. DreamGaussian: Generative Gaussian Splatting for Efficient 3D Content Creation. *The Twelfth International Conference on Learning Representations, ICLR 2024*, Vienna, Austria, May 7-11, 2024, *arXiv*, 2024. DOI: 10.48550/arXiv.2309.16653.

18. Lee, H.-H., Savva, M., & Chang, A. X. Text-to-3D Shape Generation. *arXiv*, 2024. Available from: https://arxiv.org/abs/2403.13289 (accessed 31.05.2025).

19. Chen, L., Wang, Z., Zhou, Z., Gao, T., Su, H., Zhu, J., & Li, C. MicroDreamer: Zero-shot 3D Generation in ~20 Seconds by Score-based Iterative Reconstruction, *CoRR. arXiv*, 2024. Available from: https://arxiv.org/abs/2404.19525 (accessed 31.05.2025).

20. Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., & Yang, J. Structured 3D Latents for Scalable and Versatile 3D Generation, *CoRR*, 2024, Available from: 10.48550/arXiv.2412.01506.

21. Chen, R., Chen, Y., Jiao, N., & Jia, K. Fantasia3D: Disentangling Geometry and Appearance for High-quality Text-to-3D Content Creation. *In:*

Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 22246–22256. DOI: 10.48550/arXiv.2303.1387.

22. Qiu, L., Chen, G., Gu, X., Zuo, Q., Xu, M., Wu, Y., Yuan, W., Dong, Z., Bo, L., & Han, X. Richdreamer: A generalizable normal-depth diffusion model for detail richness in text-to-3d. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 9914–9925. DOI: 10.48550/arXiv.2311.16918.

23. Li, J., Tan, H., Zhang, K., Xu, Z., Luan, F., Xu, Y., Hong, Y., Sunkavalli, K., Shakhnarovich, G., & Bi, S. Instant3D: Fast Text-to-3D with Sparse-View Generation and Large Reconstruction Model. *The Twelfth International Conference on Learning Representations, ICLR* 2024, *arXiv*. DOI: 10.48550/arXiv.2311.06214.

24. Li, M., Zhou, P., Liu, Keppo, J., Lin, M., Yan, S., & Xu, X. Instant3D: Instant Text-to-3D Generation. *International Journal of Computer Vision*, 2024, vol. 132, pp. 4456–4472. DOI: 10.1007/s11263-024-02097-5.

25. Chen, C., Yang, X., Yang, F., Feng, C., Fu, Z., Foo, C.-S., Lin, G., & Liu, F. Sculpt3D: Multi-View Consistent Text-to-3D Generation with Sparse 3D Prior. *EEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 10228–10237. DOI: 10.1109/cvpr52733.2024.00974.

26. Ren, X., Huang, J., Zeng, X., Museth, K., Fidler, S., & Williams, F. XCube: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *IEEE*, 2024, pp. 4209–4219. DOI: 10.1109/CVPR52733.2024.00403.

27. Huang, T., Zeng, Y., Zhang, Z., Xu, W., Xu, H., Xu, S., Lau, R. W., & Zuo, W. Dreamcontrol: Control-based text-to-3d generation with 3D self-prior. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5364–5373. DOI: 10.1109/CVPR52733.2024.00513.

28. Zhang, L., & et al. CLAY: A Controllable Large-scale Generative Model for Creating High-quality 3D Assets. *ACM Transactions on Graphics*, 2024, vol. no. 4, pp. 1–20. DOI: 10.1145/3658146.

29. Hyper3d.ai. *Hyper3D*. Available at: https://hyper3d.ai/ (accessed 30.05.2025).

30. Zhao, Z., & et al. Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. *arXiv*, 2025. Available at: http://arxiv.org/abs/2501.12202. (accessed 30.05.2025).

31. Zhao, Z., Liu, W., Chen, X., Zeng, X., Wang, R., Cheng, P., FU, B., Chen, T., YU, G., & Gao, S. Michelangelo: Conditional 3D Shape Generation based on Shape-Image-Text Aligned Latent Representation. *In: Thirty-seventh Conference on Neural Information Processing Systems*, *arXiv*, 2023. DOI: 10.48550/arXiv.2306.17115.

32. Li, W., Liu, J., Yan, H., Chen, R., Liang, Y., Chen, X., Tan, P., & Long, X. CraftsMan3D: High-fidelity Mesh Generation with 3D Native Generation and Interactive Geometry Refiner. *arXiv*, 2024. DOI: 10.48550/arXiv.2405.14979.

33. Cheng, Y.-C., & et al. SDFusion: Multimodal 3D Shape Completion, Reconstruction, and Generation. *In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. *IEEE*, 2023, pp. 4456–4465. DOI: 10.1109/cvpr52729.2023.00433.

34. Leng, Z., Birdal, T., Liang, X., & Tombari, F. HyperSDFusion: Bridging Hierarchical Structures in Language and Geometry for Enhanced 3D Text2Shape Generation. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 19691-19700. DOI: 10.1109/CVPR52733.2024.01862.

35. Lin, F., Hou, S., Liu, H., Gao, S., Yamada, K. D., Zhang, H. K., & Zhang, Z. Hyperbolic Chamfer Distance for Point Cloud Completion and Beyond. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2024, pp. 14549–14560. DOI: 10.1109/ICCV51070.2023.01342.

36. Koffeman, A. *Wooden Bookcase by Akoffeman (3D model)*. 2020. Available from: https://www.thingiverse.com/thing:4500930 (accessed 30.05.2025).

37. *Replica Heerenhuis Rapture Coffee Table Model, Black - Poliigon. Poliigon (3D model).* Available at: https://www.poliigon.com/model/replica-heerenhuis-rapture-coffee-table-model-black/4285 (accessed 30.05.2025).

38. Hultgren, K. *Three 1:24 Windsor Chairs. (3D Model).* Available at: https://www.thingiverse.com/ thing:21999 (accessed 30.05.2025)

Received 17.03.2025, Accepted 25.08.2025

ВПЛИВ ФОРМАТУ ПОДАННЯ ЧИСЕЛ У ПАРАМЕТРИЗОВАНИХ ТЕКСТОВИХ ЗАПИТАХ НА ТОЧНІСТЬ ГЕНЕРАЦІЇ 3D-МОДЕЛЕЙ

О. Ю. Барковська, Л. Д. Бухарова, І. В. Рубан, О. С. Ляшенко, А. А. Коваленко, В. С. Сердечний

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Предметом вивчення в статті є аналіз моделей генерації 3D-об'єктів на основі текстового запиту із різними форматами подання числових параметрів. **Метою роботи** є оцінка та порівняння ефективності моделей Hunyuan3D 2.0, Michelangelo та SDFusion в залежності від формату вхідного текстового запиту за метриками часу генерації, CLIP-Similarity та Chamfer Distance. **Для досягнення мети** були **поставлені наступні завдання**: 1) провести систематичний огляд сучасних генеративних 3D-моделей щодо їх здатності обробляти запити з числовими параметрами, поданими в різних формах; 2) розробити методику формування вхідних текстових запитів з числовими описами в різних форматах; 3) провести кількісну та якісну оцінку генеративних 3D моделей за метриками часу генерації, CLIP-Similarity та Chamfer Distance, враховуючи візуальну подібність текстовим описам та геометричну подібність до Ground Truth об'єктів; 4) проаналізувати отримані результати оцінки генеративних моделей текст-в-3D для завдань параметризованої генерації 3D-об'єктів для подальшої імплементації в прикладних користувацьких інтерфейсах. **Використаними в даній роботі методами є**: методи машинного навчання, методи векторного представлення тексту та зображень, статистичні методи оцінки результатів та евристичний метод формування текстових запитів. **Отримані результати** показують, що не існує універсальної генеративної 3D моделі, яка спроможна створювати об'єкти, що повністю відповідають параметризованому текстовому запиту. Запропонована методика дозволяє формувати вхідні текстові послідовності, що містять числові представлення включаючи: десяткові дроби, словесні представлення або звичайні дроби, для проведення подальшого аналізу точності генерування об'єктів. **Висновки**. Підтверджено подібну ефективність навчання 3D генеративних моделей у спільному латентному просторі текстових ознак, що містять числові параметри та використання датасетів із чітко визначеними геометричними характеристиками об'єктів. Результати показують, що модель Hunuyan3D-2.0 придатна до подальшого дослідження та модифікації з метою адаптації використовуваних методів для створення персоналізованих тривимірних об'єктів за заданими числовими параметрами, таких як накладка на протез.

Ключові слова: GenAI; подібність; точність; параметри; Chamfer Distance; CLIP-Similarity; текст-в-3D.

Барковська Олеся Юріївна – канд. техн. наук, доц., доц. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Бухарова Любов Дмитрівна – маг. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Рубан Ігор Вікторович –д-р техн. наук, проф., проф. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Ляшенко Олексій Сергійович - канд. техн. наук, доц., доц. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

Коваленко Андрій Анатолійович –д-р техн. наук, проф., проф. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Сердечний Віталій Сергійович - асп. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна.

Olesia Barkovska - PhD, Associate Professor, Associate Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: olesia.barkovska@nure.ua, ORCID: 0000-0001-7496-4353, Scopus Author ID: 24482907700.

Liubov Bukharova – Master Student of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: liubov.bukharova@nure.ua, ORCID: 0009-0003-0635-8695.

Igor Ruban – Doctor of Technical Sciences, Professor, Professor of Department Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: ihor.ruban@nure.ua, ORCID: 0000-0002-4738-3286, Scopus Author ID: 7004018101.

Oleksii Liashenko – PhD, Associate Professor, Associate Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: oleksii.liashenko@nure.ua, ORCID: 0000-0002-0146-3934, Scopus Author ID: 55658561300.

Andriy Kovalenko – Doctor of Technical Sciences, Professor, Professor of Department Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: andriy.kovalenko@nure.ua, ORCID: 0000-0002-2817-9036, Scopus Author ID: 56423229200.

Vitalii Serdechnyi – PhD Student of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: vitalii.serdechnyi@nure.ua, ORCID: 0009-0007-8828-5803, Scopus Author ID: 57210288841.