

Generative Design

Final project, GenAI

Authors:
Anastasia Volkova,
Maksim Osipenko,
Iana Kulichenko,
Maksim Komiakov



1. I-Design: Personalized LLM Interior Designer

Paper: <https://arxiv.org/abs/2404.02838>



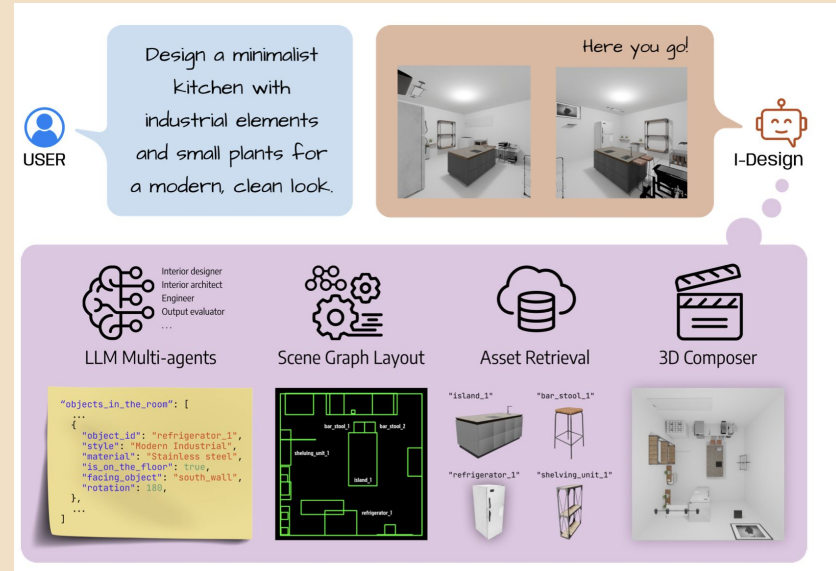
Task: Creation of a 3D interior scene based on unstructured text input from the user T_{user} .

The room dimensions are specified as $(l_{\text{room}}, w_{\text{room}}, h_{\text{room}}) \in \mathbb{R}^3$, and the number of objects $n \in \mathbb{R}$ that should be included in the scene. The goal is to create a 3D scene that meets the user's requests.

The scene is described by a graph $G = (O, E)$, where O is the set of objects and E represents the relationships between them.

Workflow Overview

1. **LLM Agents:** Convert user input into a scene graph (objects + relationships).
2. **Backtracking Algorithm:** Places objects hierarchically, starting from root nodes.
3. **3D Asset Retrieval:** Generates text descriptions for each object, uses CLIP to match with 3D assets.
4. **3D Rendering:** Visualizes the scene with a renderer.
5. **Final Output:** Includes scene graph, floor plan, preset viewpoints, and interaction logs.



Agents

I-Design: Personalized LLM Interior Designer Agents

7

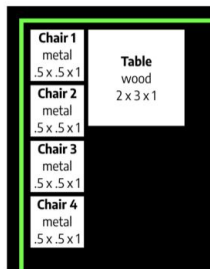
User prompt

Design specifications in natural language

Design a kitchen with a table and four chairs

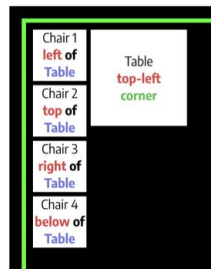
Interior Designer

Proposes named objects, sizes, styles



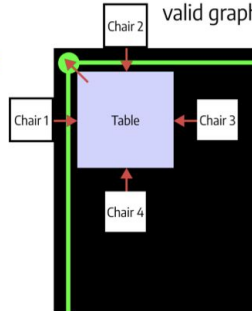
Interior Architect

Proposes pairwise object relations



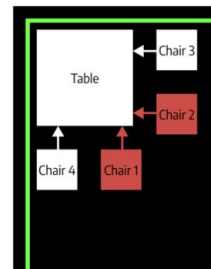
Engineer

Creates and ensures a valid graph



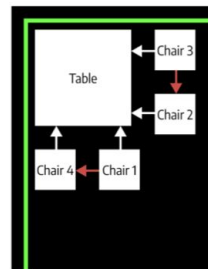
Layout Corrector

Fixes invalid and implausible relations



Layout Refiner

Introduces relations into object clusters



Interior Designer:

Suggests a set of objects for the scene and their properties, including name, material, dimensions, orientation, and position.

Interior Architect:

Defines the relationships between objects and their connection to the room layout.

Engineer:

Constructs and converts the relative scene graph into a structured JSON object

Layout Corrector:

Fixes invalid connections in the graph, such as objects exceeding room boundaries or incompatible relationships between objects.

Layout Refiner:

Resolves ambiguities between child nodes that share the same relationship with the parent object.

Our experiments

We used the models llama-3.2-11b-text-preview, llama3-8b-8192, and gpt-3.5-turbo-1106 for agent communication.

We successfully launched the agents' interaction and the proposed arrangement of objects in the room, but the scene graph construction had errors in the authors' code, and the JSON file formats were breaking. As a result, we decided to continue experimenting with a change in strategy.

Admin (to chat_manager):

```
The room has the size 4.0m x 4.0m x 2.5m.
```

```
User Preference:
```

```
```
```

```
A creative vibrant livingroom
```

```
```
```

```
Room layout elements:
```

```
```
```

```
['south_wall', 'north_wall', 'west_wall', 'east_wall', 'middle of the room', 'ceiling']
```

```
```
```

```
json
```

Our experiments

Interior_designer (to chat_manager):

Based on the user's preference for a creative and vibrant living room, I suggest the following layout and objects.

Room Layout

- * South Wall:
 - * Object 1: TV Stand (Style: Modern, Material: Glass and Metal, Dimensions: length=1.5m, width=0.8m, height=0.8m, Quantity: 1)
 - * Object 2: Decorative Shelf (Style: Modern, Material: Wood and Glass, Dimensions: length=1.5m, width=0.4m, height=0.5m, Quantity: 1)
- * North Wall:
 - * Object 3: Large Sofa (Style: Modern, Material: Leather, Dimensions: length=3.5m, width=1.5m, height=0.9m, Quantity: 1)
 - * Object 4: Side Table (Style: Modern, Material: Glass and Metal, Dimensions: length=0.8m, width=0.6m, height=0.8m, Quantity: 2)
- * West Wall:
 - * Object 5: Large Window (Style: Modern, Material: Wood and Glass, Dimensions: length=3.0m, width=1.0m, height=2.5m, Quantity: 1)
 - * Object 6: Floor Lamp (Style: Modern, Material: Glass and Metal, Dimensions: length=1.0m, width=0.5m, height=1.5m, Quantity: 1)
- * East Wall:
 - * Object 7: Decorative Corner (Style: Modern, Material: Wood and Glass, Dimensions: length=1.0m, width=1.0m, height=1.5m, Quantity: 1)
 - * Object 8: TV (Style: Modern, Material: Glass and Metal, Dimensions: length=1.5m, width=1.0m, height=0.8m, Quantity: 1)
- * Middle of the room:
 - * Object 9: Coffee Table (Style: Modern, Material: Glass and Metal, Dimensions: length=1.5m, width=0.8m, height=0.4m, Quantity: 1)
 - * Object 10: Plants (Style: Modern, Material: Ceramic and Wood, Dimensions: length=0.5m, width=0.5m, height=1.5m, Quantity: 3)
- * Ceiling:
 - * Object 11: Ceiling Light (Style: Modern, Material: Glass and Metal, Dimensions: length=1.0m, width=1.0m, height=1.0m, Quantity: 4)

Our experiments

Next speaker: Interior_architect

Interior_architect (to chat_manager):

Based on the user's preference for a creative and vibrant living room, I analyze the provided layout and objects and suggest optimal placement for each object:

Suggested Placement for Each Object:

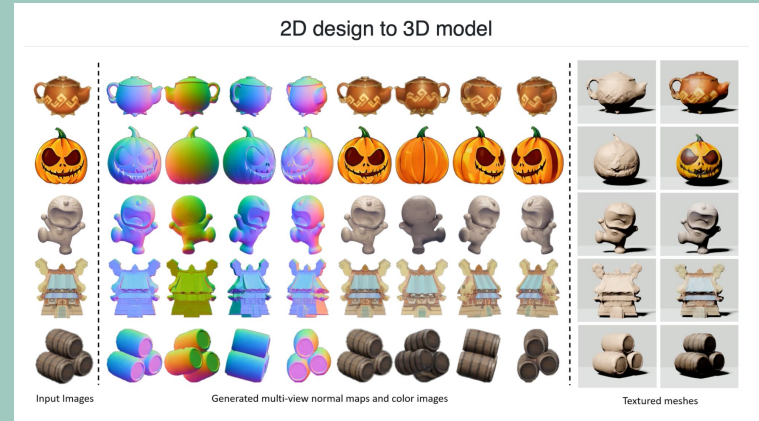
1. TV Stand:
 - * Placement: In the south wall
 - * Proximity: Adjacent to the TV
 - * Facing: Facing the TV
2. Decorative Shelf:
 - * Placement: On the south wall
 - * Proximity: Above the TV Stand
 - * Facing: Facing the north wall
3. Large Sofa:
 - * Placement: On the north wall
 - * Proximity: Not adjacent to any other objects
 - * Facing: Facing the south wall
4. Side Table:
 - * Placement: Adjacent to the Large Sofa
 - * Proximity: On the same wall as the Large Sofa
 - * Facing: Facing the Large Sofa
5. Large Window:
 - * Placement: On the west wall
 - * Proximity: Not adjacent to any other objects
 - * Facing: Facing the east wall
6. Floor Lamp:
 - * Placement: In the middle of the room
 - * Proximity: Adjacent to the Large Sofa
 - * Facing: Facing the Large Sofa
7. Decorative Corner:
 - * Placement: In the east wall
 - * Proximity: Not adjacent to any other objects
 - * Facing: Facing the west wall
8. TV:
 - * Placement: On the east wall
 - * Proximity: Not adjacent to any other objects
 - * Facing: Facing the Large Sofa
9. Coffee Table:
 - * Placement: In the middle of the room
 - * Proximity: Adjacent to the Large Sofa
 - * Facing: Facing the Large Sofa
10. Plants:
 - * Placement: Around the Large Sofa
 - * Proximity: Adjacent to the Large Sofa
 - * Facing: Facing the Large Sofa
11. Ceiling Light:
 - * Placement: On the ceiling
 - * Proximity: Not adjacent to any other objects
 - * Facing: Facing the ceiling

2. Text->Image Image->3D Using DeepFloyd IF and Wonder3D

Papers:

<https://arxiv.org/pdf/2205.11487>

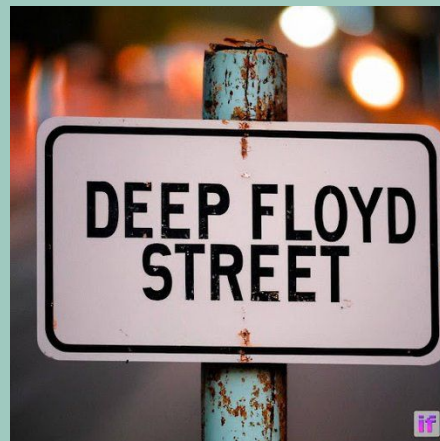
<https://arxiv.org/abs/2310.15008>



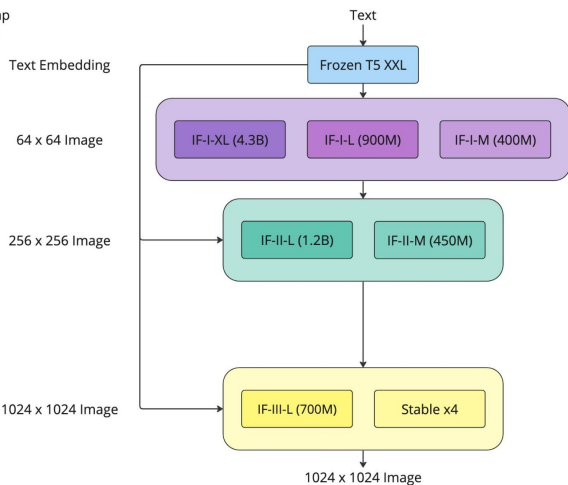
We are generating a 2D image of the apartment layout (or a section of the apartment's furniture) from the text → then converting this 2D image (either of the apartment layout or furniture section) into a mesh. For text-to-image generation we use DeepFloyd IF model and for image to 3D we use Wonder 3D.

IF by DeepFloyd: IF-I-XL-v1.0

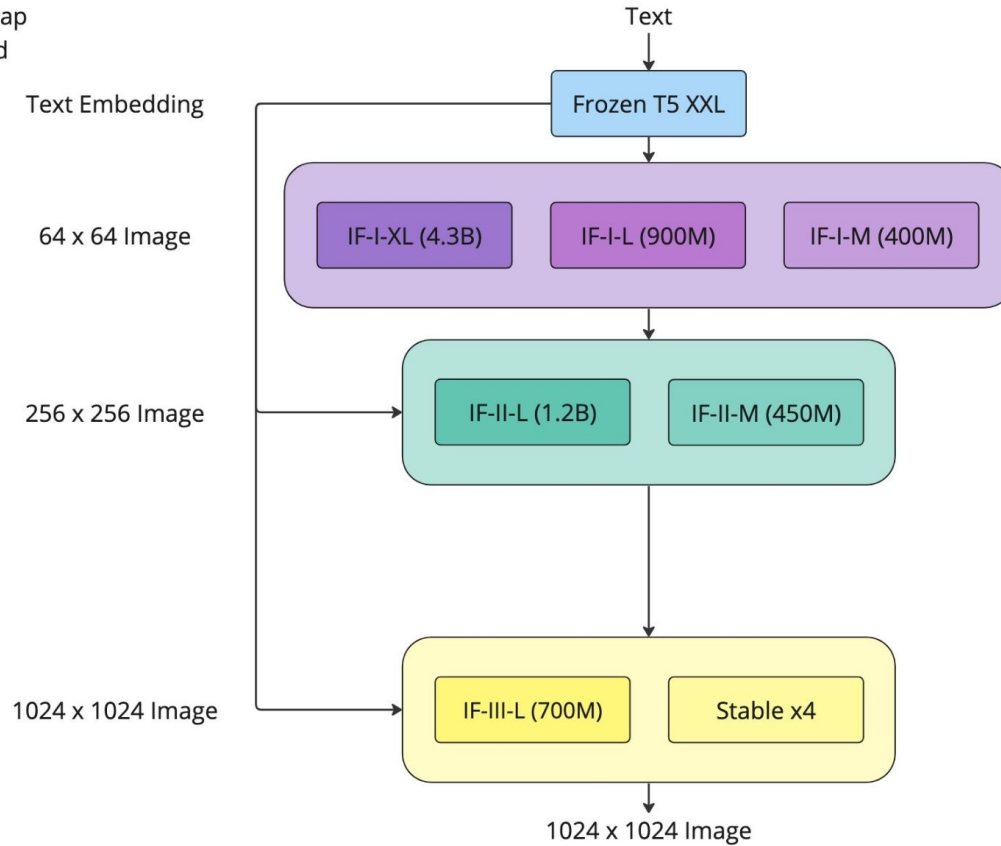
Text-to-image open source
cascaded pixel diffusion
model



a photo of a violet baseball cap
with yellow text "deep floyd
better than text"



a photo of a violet baseball cap
with yellow text "deep floyd
better than text"



Text to Image

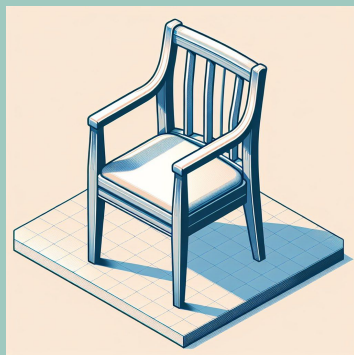
**Text
prompt**

The chair with the
white background

The chair in axonometric
projection without the
background

A wooden chair with a dark brown finish. The chair has a solid, flat seat and four straight legs. Its back features a simple, elegant design with two diagonal crossbars forming an "X" shape, supported by vertical and horizontal bars. The chair has a minimalist, classic appearance, suitable for dining or casual seating. The background is a plain white, enhancing the chair's visual prominence.

Dalle



Text to Image with IF

Text prompt

A wooden chair with a dark brown finish. The chair has a solid, flat seat and four straight legs. Its back features a simple, elegant design with two diagonal crossbars forming an "X" shape, supported by vertical and horizontal bars. The chair has a minimalist, classic appearance, suitable for dining or casual seating. The background is a plain white, enhancing the chair's visual prominence.



Create an image of a modern three-seater sofa with a minimalist design. It has a structured, rectangular shape with clean lines and sharp edges. The sofa is upholstered in a smooth, grey fabric. The backrest consists of three separate cushions, matching the darker grey seat cushions. The armrests are wide and slightly lower than the backrest. The sofa is elevated on four slim, angled legs, giving it a mid-century modern look. The background is plain light grey, emphasizing the sofa's design.

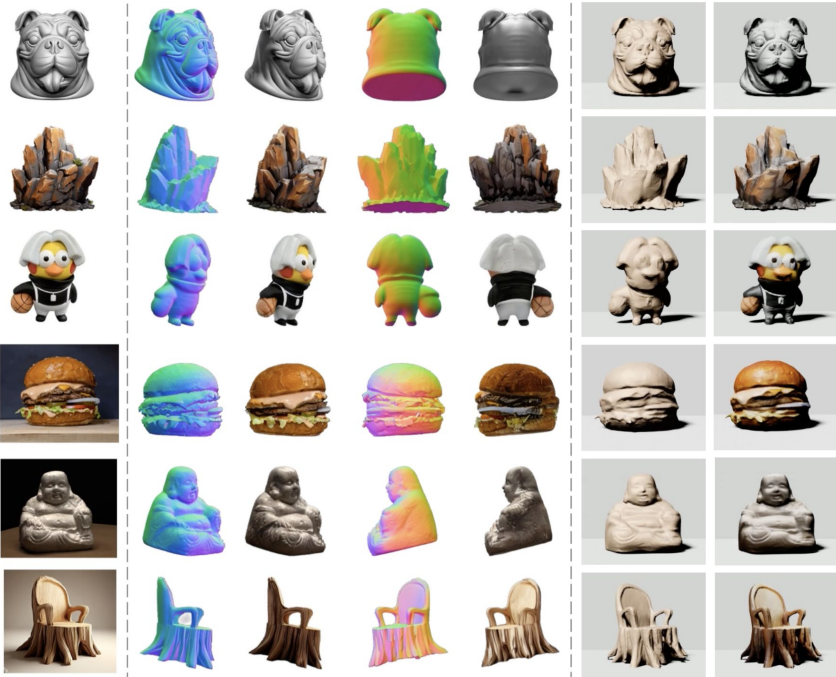


Create a picture of a tall, wooden wardrobe with a minimalist design. It features two large doors on the top section, each with a recessed panel, creating a simple and clean look. Below the doors are two horizontal drawers, each with two small round handles. The wardrobe has a flat top with a slight overhang and a flat base, adding to its sleek and contemporary appearance. The handles and knobs are metallic, contrasting subtly with the matte grey finish of the wardrobe.



IF

Wonder3D produces consistent multi-view normal maps and corresponding color images, and thus reconstructs high-fidelity textured mesh from a single image in only 2~3 minutes.



Input Images

Generated multi-view normal maps & color images

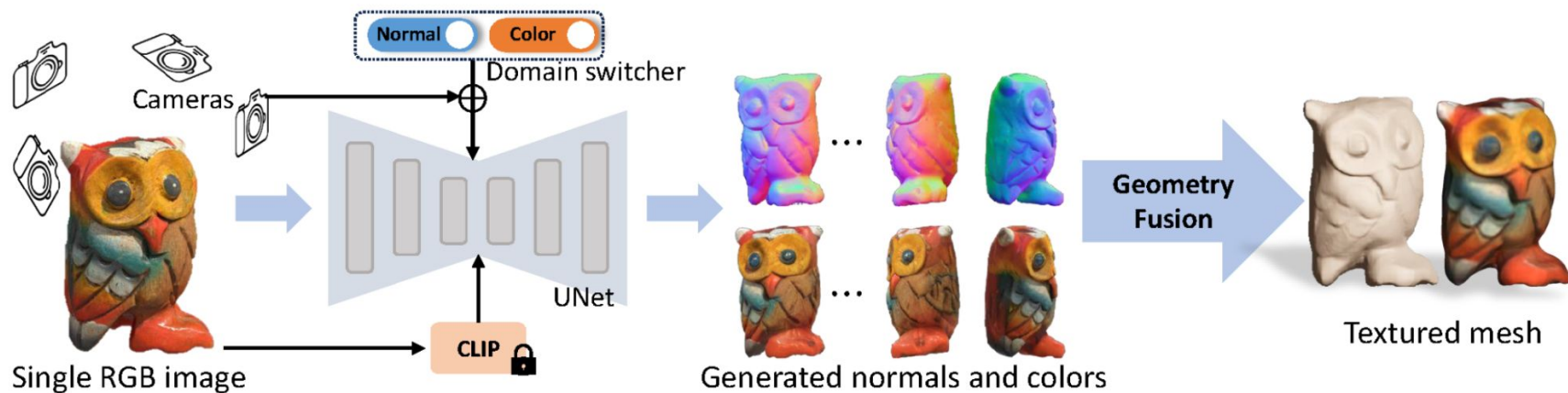
Textured meshes

Wonder3D for image to 3D

Paper: <https://arxiv.org/abs/2310.15008>

Method for efficiently generating high-fidelity
textured meshes from single-view images

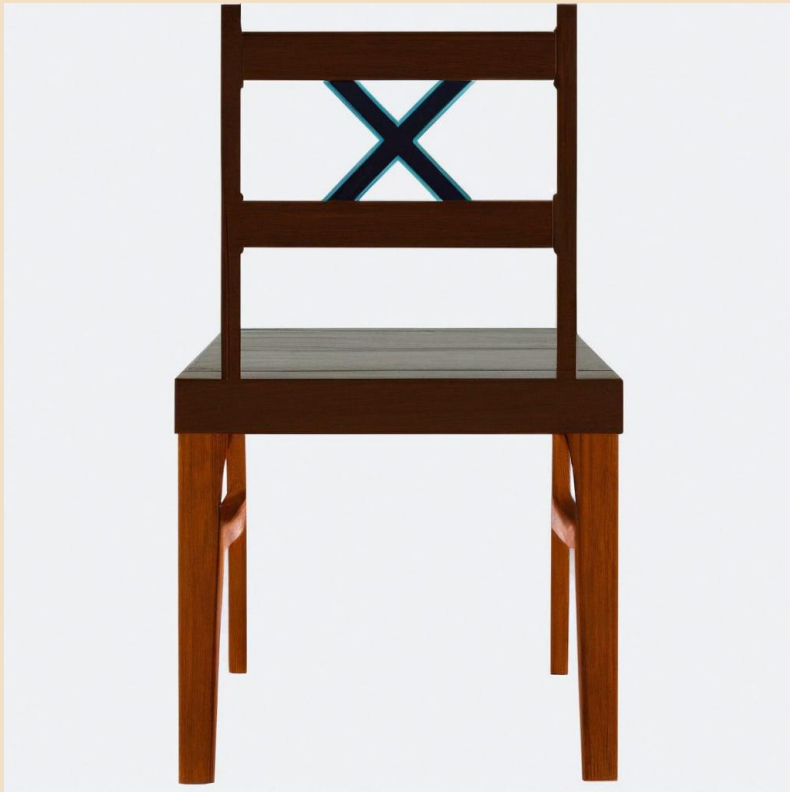
Overview of Wonder3D



Overview of Wonder3D.

Given a single image, Wonder3D takes the input image, the text embedding produced by CLIP model, the camera parameters of multiple views, and a domain switcher as conditioning to generate consistent multi-view normal maps and color images. Subsequently, Wonder3D employs an innovative normal fusion algorithm to robustly reconstruct high-quality 3D geometry from the 2D representations, yielding high-fidelity textured meshes.

Our experiments

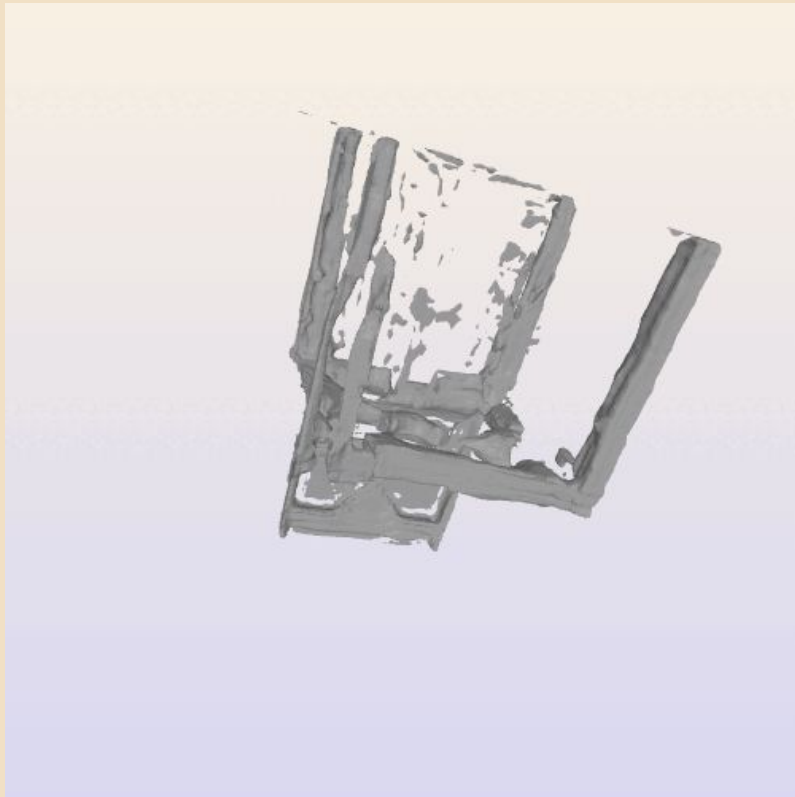


Chair image generated



Sofa image generated

Our experiments



Chair mesh



Sofa mesh

	Metrics			
	Chair	Sofa	Wardrobe	
	SSIM: 0.7591	SSIM: 0.7756	SSIM: 0.7906	
	MSE: 0.1328	MSE: 0.1277	MSE: 0.0462	
	LPIPS: 0.5937	LPIPS: 0.5543	LPIPS: 0.5163	

Metrics

Text to 2d Image Human-in-the-Loop Feedback: Involving the Experts

We asked GPT-4o to rate the performance of the generation based on prompt and image



Relevance: Does the retrieved information address the query	10	7	6
Correctness: Is the content factually accurate	9	6	5
Fluency: Does it read well, or does it feel awkward or forced	10	N/A	N/A
Completeness: Does it cover the question fully or leave gaps	10	6	5
Mean:	9.75	6.33	5.33

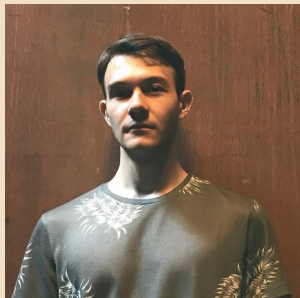
Conclusions

1. The communication system for the team of large language model agents was implemented in i-Design, although the 3D scene generation process was unsuccessful.
2. The entire process of the “text prompt to 3d mesh of the room” generation was divided into two distinct phases: the first involved transforming the text into a 2d image (IF), while the second focused on converting the 2d image into a 3d mesh (Wonder3D).
3. Room design process performance failed because of background mesh generation issues.
4. Interior objects generation from text to 3d mesh was implemented.
5. Due to the resulting metrics, the performance may be evaluated as satisfactory.

Future Plans

1. **Redefine 3d scene generation in i-Design**
2. **Score the performance of text-to-3dmesh generation with the ground truth labels**
3. **Add multimodality to model (not only to metric measuring)**

Our team



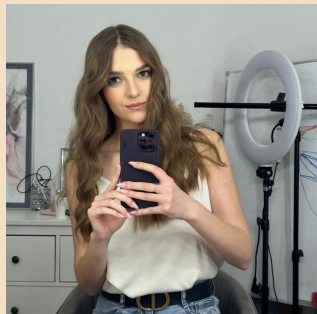
Maksim Osipenko:

worked on image to mesh task,
Wonder3D and mesh creation



Anastasia Volkova:

worked on text to image IF by
DeepFloyd, literature overview



Iana Kulichenko:

worked on implementation of
IDesign, presentation



Maksim Komiako:

worked on metrics calculation,
reports, presentation

Thanks

Generative Design

Authors:

Anastasia Volkova,

Maksim Osipenko,

Iana Kulichenko,

Maksim Komiakov



Appendix

Image-3dmesh of interior

