

# Harnessing Layered Graphic Designs with Real Intentions for Text-to-Design Generation

## Supplementary Material

### A. Dataset Details

**Samples.** Figure 1 shows more samples from our dataset.

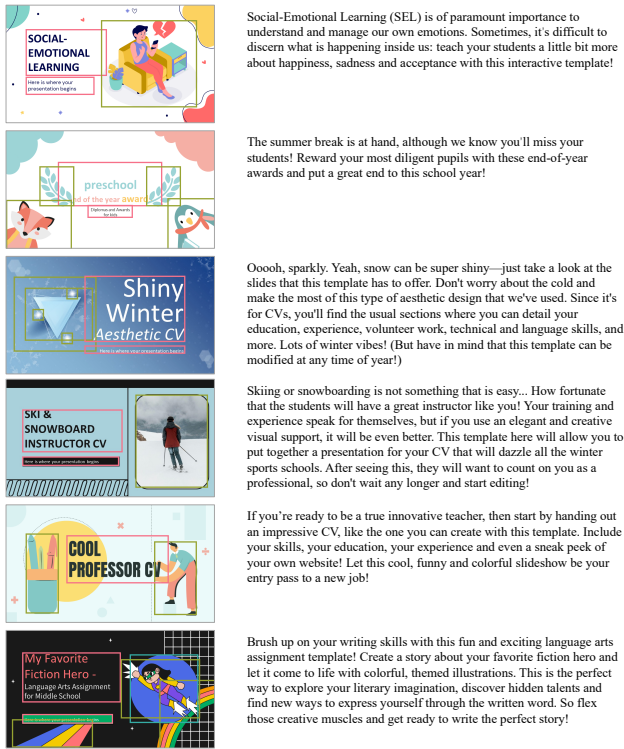


Figure 1. Samples in the LADEREIN dataset.

**Tag Distribution.** Figure 2 illustrates the distribution of tags in our dataset.

### B. Text Content Generation Prompt

Figure 3 shows the prompt used to generate text contents for design rendering.

### C. User Study Details

We present the details of the user study for comparing real and synthetic design intentions in Figure 5.

### D. More Evaluation Results of Our Dataset

**Visual Comparison of DesignDiff-Syn and DesignDiff-Real.** We show some designs generated by DesignDiff-Syn and DesignDiff-Real in Figure 6.

**Evaluation of SDXL-Crello and SDXL-LADEREIN on Augmented Prompts.** We utilize GPT to augment design intentions from Crello and LADEREIN into long, detailed

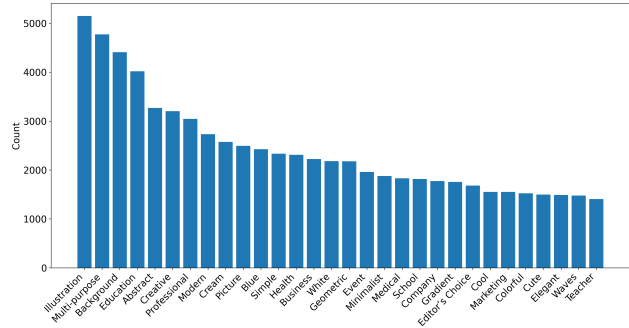


Figure 2. Distribution of the top 30 Tags in the LADEREIN dataset. The horizontal axis represents tag name, while the vertical axis indicates tag count.

You are a top-tier slide content strategist and copywriter. Based on the user-provided [Design Intention] and [Specified Word Count], generate ready-to-use content for a slide textbox. You must use the [Specified Word Count] as the primary cue to determine the text's "role" and strictly write in that role's style: When the [Specified Word Count] is short (e.g., 3-7 words): Your inferred role: This is a [Main Title], [Section Header], or [Key Point] Your writing style: The content must be extremely concise, punchy, impactful, and memorable. When the [Specified Word Count] is long (e.g., 8-15 words): Your inferred role: This is a [Subtitle] or a [Context-Setting Header]. Your writing style: The content must complement a main idea (the title), not just summarize the [Design Intention]. It should provide immediate context, clarify the scope of the topic, or pose the key question the slide will answer. It must be clear, concise, and serve as a supportive secondary line to a main headline. Your output must ABSOLUTELY NOT contain any analysis of your "role inference," explanations, or other pleasantries. Output only one piece of copy that meets the requirements.

Figure 3. Prompt for generating text contents for design rendering. The word count for a text element is calculated based on its predicted font size and bounding box size.

prompts, based on the prompt shown in in Figure 4. Table 1 presents the quantitative results of SDXL-Crello and SDXL-LADEREIN on the augmented prompts. SDXL-LADEREIN outperforms SDXL-Crello across all the resolutions and test sets except for the  $256 \times 512$  resolution where SDXL-LADEREIN achieves slightly lower DCLIP scores on the two test sets.

You are an expert prompt engineer for SDXL. Your task is to turn a Design Intention into a Caption for SDXL image generation. This caption must be detailed and full of visual words. But it shouldn't be too long. Here is an example for you to learn from:

Design Intention: Create a birthday card with a festive and colorful design featuring a cake with fruits. The card is an invitation to a birthday party happening on June 22nd from 8pm to 11pm at 123 st. Any city 4567.

Caption: A burst of vibrant pink hues dominates the canvas: a mix of blush, fuchsia and rose clothing neatly arranged. In the corner, a gleaming starburst accentuates a glowing testimonial from Jenny Wilson, her words of praise shimmering. Beneath the testimonial, elegantly scripted contact information is highlighted: phone number 123-456-789, website WWW.FASHIONSTYLE.COM, and an address at 123 Anywhere ST., Any City.

Figure 4. Prompt for generating augmented prompts.

Table 1. Quantitative results of fine-tuning SDXL on the Crello and LADEREIN training sets (with augmented prompts), and testing them on the test splits (with augmented prompts) of the two datasets. For each resolution, the two models are only allowed to generate images of the resolution for computing the quantitative scores.

Resolution	Model	Crello		LADEREIN	
		FID↓	DCLIP↑	FID↓	DCLIP↑
1024 × 1024	SDXL-Crello	12.78	20.27	19.81	16.94
	SDXL-LADEREIN	<b>9.382</b>	<b>20.50</b>	<b>11.22</b>	<b>17.17</b>
960 × 512	SDXL-Crello	12.65	19.98	14.52	16.88
	SDXL-LADEREIN	<b>7.953</b>	<b>20.25</b>	<b>9.078</b>	<b>17.09</b>
256 × 512	SDXL-Crello	9.890	<b>19.60</b>	10.87	<b>16.58</b>
	SDXL-LADEREIN	<b>6.289</b>	19.00	<b>9.871</b>	16.55

## E. More Evaluation Results of DesignDiff

We show the designs generated by different models at the 1024 × 1024 resolution in Figure 7.

Welcome to our user study on design intention analysis. In this study, you will be presented with 30 rounds of evaluation tasks.

In each round, you will see:

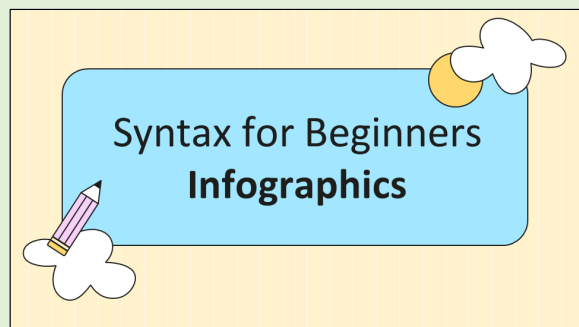
1. A design template image
2. Two different design intentions (Intention A and Intention B)

Your task is to evaluate which intention better reflects the true design purpose of the template, focusing solely on the content and meaning of the intentions, regardless of their wording or sentence structure.

Important guidelines:

- Consider the design template's visual elements, layout, and overall composition
- Focus on the core message and purpose that the design template aims to convey
- Compare the two intentions based on their content and meaning, not their wording
- There are no right or wrong answers - we are interested in your honest assessment
- Take your time to carefully evaluate each round

Please proceed to the first round when you are ready. You can take breaks between rounds if needed.



- Intention A: Is syntax something that sounds difficult to you? OK, let's see, in English, all sentences follow the structure "subject-verb-object". Let's see an example: "You download templates". "You" is the subject, "download" is the verb, and "templates" is the object. Easy! By the way, we mean it: you can download this template. It's full of funny and colorful infographics that you can use to teach syntax for beginners (or young learners). Remember that it's just a template, so the contents and the actual examples is something you need to add yourself!"
- Intention B: Create a PowerPoint slide aimed at beginners in education, particularly focusing on teaching syntax and grammar through cute and visually appealing infographics that can enhance the learning experience.

Figure 5. User study details. The instruction (left) and one example question (right) are shown.

Design Intention	DesignDiff-Syn	DesignDiff-Real
<p>Medical notes don't have to look sterile... You'd like them to be organized and understandable at a glance... it was originally engineered for medical note-taking... keep your notes so neat...</p>	<p>Transform Medical Notes Today!</p>	<p>Revolutionize Your Medical Notes!</p>
<p>Looking for a career in the food and beverage industry but need clarity... In order to entice possible people from the industry, we've created this template in an attractive way...</p>	<p>Food &amp; Beverage Manager Template</p>	<p>Crafting Your Food &amp; Beverage Career Path</p>
<p>... Our upcoming workshop on writing final projects ... by our visually stunning... so all you need to bring is your willingness to learn...</p>	<p>Transform your final projects with innovative writing techniques and stunning visual aids!</p>	<p>Join our workshop: Elevate your final projects with stunning visuals and expert guidance!</p>
<p>... Are you thinking of creating your restaurant's app... It is a colorful... quickly capture the attention of your audience...</p>	<p>Engage Your Customers</p>	<p>Create Your Restaurant App!</p>

Figure 6. Designs generated by DesignDiff-Syn and DesignDiff-Real given the input design intentions (on the leftmost). In each intention, we highlight the design purpose (orange), feelings to convey (blue), and expected audience actions (green).

Design Intention	SDXL	OpenCOLE	DesignAR (Ours)	DesignDiff (Ours)
<p>... If you are in love with 80's clothing and want to launch your own brand,... to develop your marketing plan. It has a creative retro style... bring a jovial and fun touch...</p>				
<p>This lesson is the perfect way to have fun while learning about... proving triangle congruency... to help make math more fun and approachable.</p>				
<p>Maintaining a healthy relationship... is the key to living happily. A workshop on how to do this... giving it a more distinctive personality. Communicate, talk with your students, reach them...</p>				
<p>..."Weather Vocabulary - French - Pre-K" presentation... and create big learning experiences for the littlest students! Dynamic and adorable... makes your lessons more playful and exciting...</p>				

Figure 7. Comparison of designs at  $1024 \times 1024$  resolution generated by different models from the input design intentions (on the leftmost). In each intention, we highlight the design purpose (orange), feelings to convey (blue), and expected audience actions (green).