

DreamStruct: Understanding Slides and User Interfaces via Synthetic Data Generation

Yi-Hao Peng, Faria Huq, Yue Jiang, Jason Wu, Amanda Li, Jeffrey Bigham, Amy Pavel (ECCV 2024)



“TL;DR: Generating Declarative Language (e.g., HTML Code) with Labels for Visual Design Representation Learning”

I. Ideation

$$C_d = \mathcal{G}_D(c_s, P, L)$$

c_s : Seed design examples
 P : Design guidelines and patterns
 L : Targeted semantic labels
 \mathcal{G}_D : Description generation function
 Condition: BERTScore(c_d) < 0.7

Generated Slide Description

💡 — This slide displays the subject of mastering Japanese verb conjugations, titled 'Common Verb Conjugations' at the top. It features a bar graph quantifying the use of different verbs such as 'Taberu,' 'Kiku,' and others, in varied colors. Directly below, a descriptive table elucidates the transformation of verbs from the plain form to the more formal -masu and past tense forms, with 'Taberu' transitioning to 'Tabemasu' and 'Tabemashita'. The slide concludes with a footer of the copyright notice at the bottom."

Generated UI Description

💡 — "The UI screen presents a 'Recipe Finder' app, featuring a top search bar, filter toggles for 'Vegetarian,' 'Under 30 Minutes,' and 'Low-Calorie,' and displays two recipes: 'Creamy Pasta' (Easy, 6 ingredients) and 'Garden Salad' (Medium, 5 ingredients), each accompanied by an image. The bottom of the screen includes a copyright notice for 2023."

II. Generation & III. Production

$$c_h = \mathcal{G}_H(c_d, P, L) \quad C_{dhm} = \{\mathcal{R}(\mathcal{F}(c_h)) \mid \forall c_d \in C_d, \text{CLIP}(c_d, \mathcal{R}(\mathcal{F}(c_h))) > 0.3\}$$

c_d : Generated design concept
 P : Design guidelines and patterns
 L : Targeted semantic labels
 \mathcal{G}_H : Code generation function

c_h : Generated labeled code
 \mathcal{F} : Post-processing function (ensure interactive UI visibility)
 \mathcal{R} : Rendering function (replace visual placeholder)
 Condition: $\text{CLIP}(c_d, \mathcal{R}(\mathcal{F}(c_h))) > 0.3$

Generated Slide Code (HTML)

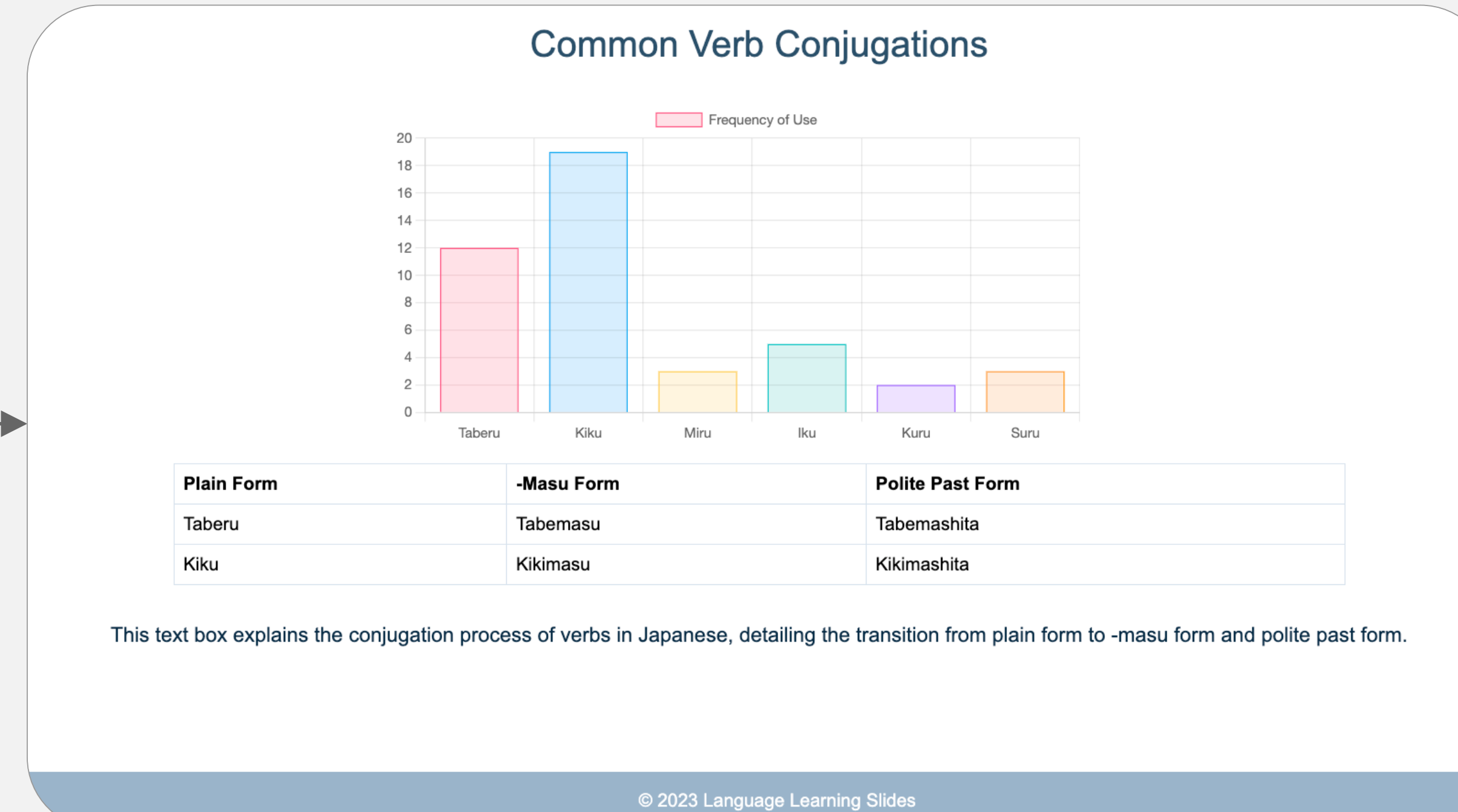
```
<!DOCTYPE html>
<html lang="en">
<head>
<style>
[...]
```

CSS
Style & Layout

HTML
Structure & Metadata

JS
Chart Elements

Generated Slide



CSS

```
.header {
font-size: 32px;
color: #334E68;
padding: 16px;
text-align: center;
}
.footer {
position: absolute;
bottom: 0;
width: 100%;
padding: 16px;
background-color: #9FB3C8;
text-align: center;
}
```

HTML

```
<div class="container">
<div class="header"><span data-type="title">Common Verb Conjugations</span></div>
<div class="chart-container">
<canvas data-type="chart" id="bar-chart"></canvas>
</div>
<div class="table-container">
<table data-type="table" class="table">
<thead>
<tr>
<th>Plain Form</th>
<th>-Masu Form</th>
<th>Polite Past Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taberu</td>
<td>Tabemasu</td>
<td>Tabemashita</td>
</tr>
<tr>
<td>Kiku</td>
<td>Kikimasu</td>
<td>Kikimashita</td>
</tr>
</tbody>
</table>
</div>
</div>
```

JS

```
var ctx = document.getElementById('bar-chart').getContext('2d');
var myChart = new Chart(ctx, {
type: 'bar',
data: {
labels: ['Taberu', 'Kiku', 'Miru', 'Iku', 'Kuru', 'Suru'],
datasets: [
label: 'Frequency of Use',
data: [12, 19, 3, 5, 2, 3],
backgroundColor: [
...
],
borderColor: [
rgb(255, 99, 132, 1);
...
]
}
});
```

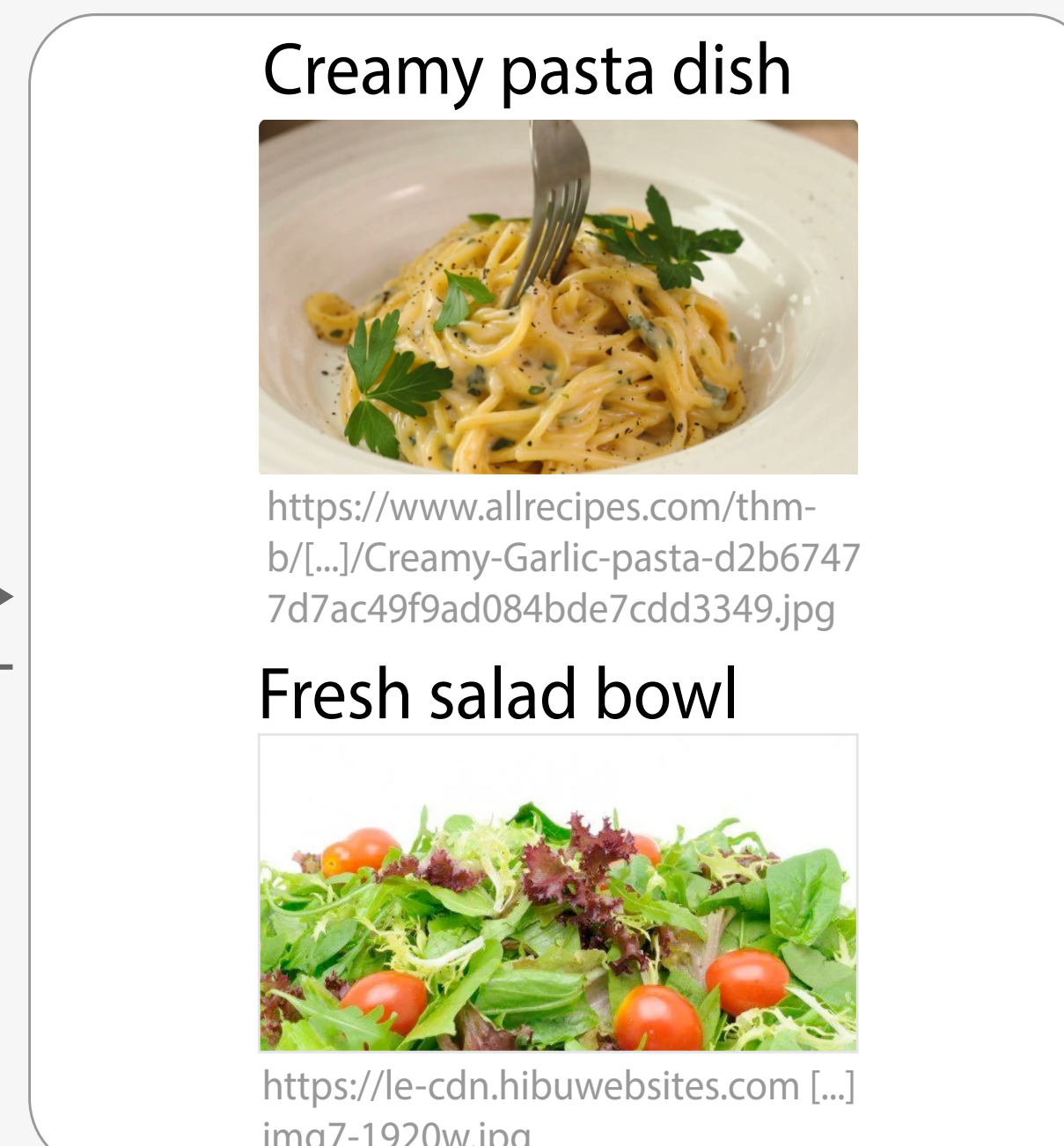
Generated UI Code (HTML)

```
<div class="recipe-card">

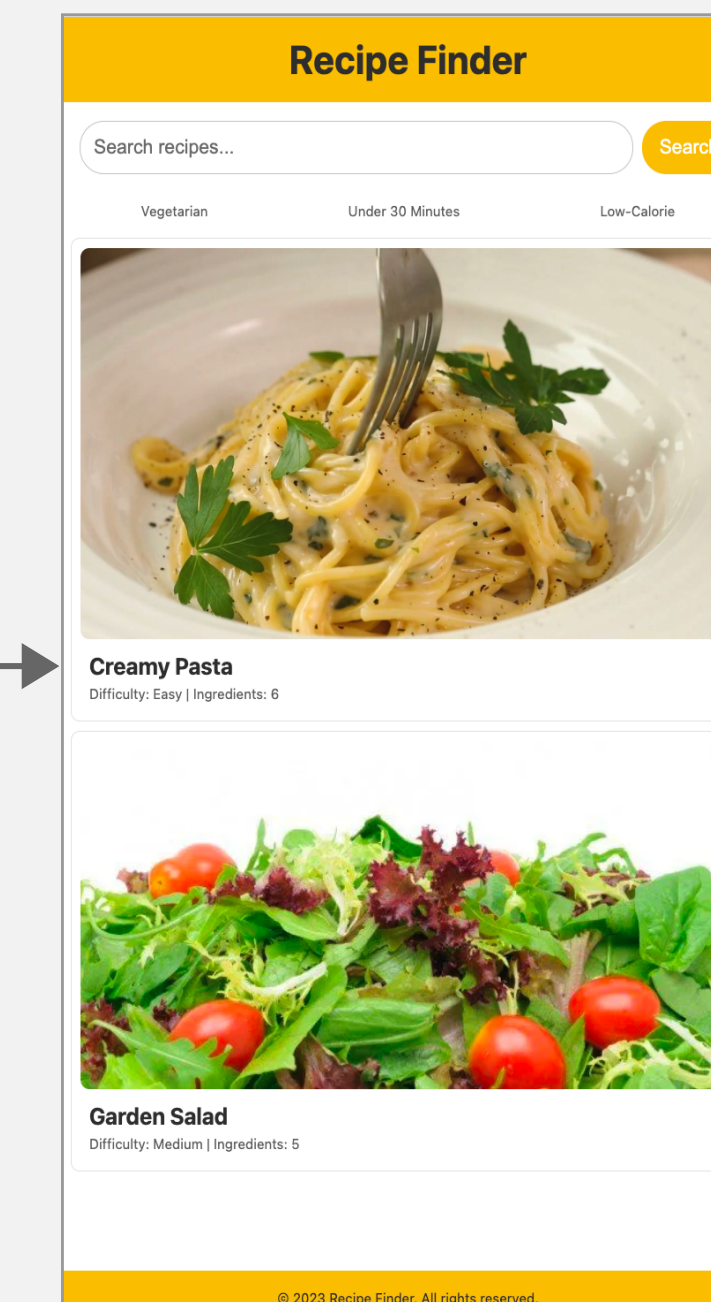
<div class="recipe-info">
<div class="recipe-title">Creamy Pasta</div>
<div class="recipe-details">Difficulty: Easy | Ingredients: 6</div>
</div>
<div class="recipe-card">

<div class="recipe-info">
<div class="recipe-title">Garden Salad</div>
<div class="recipe-details">Difficulty: Medium | Ingredients: 5</div>
</div>
</div>
```

Image Search/Generation

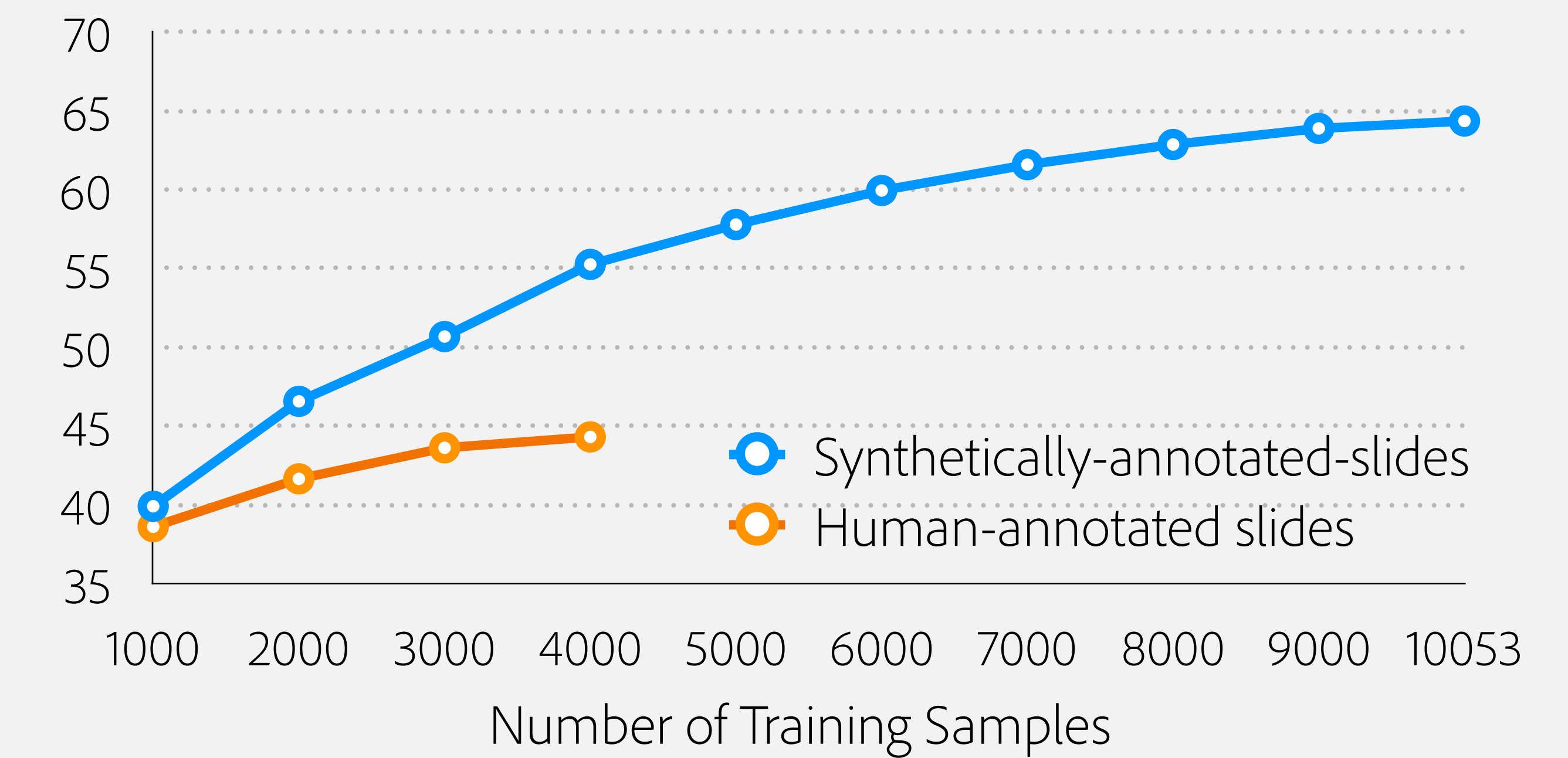


Generated UI



Results

Slide Element Recognition (mAP%)



UI Element Recognition (mAP%)

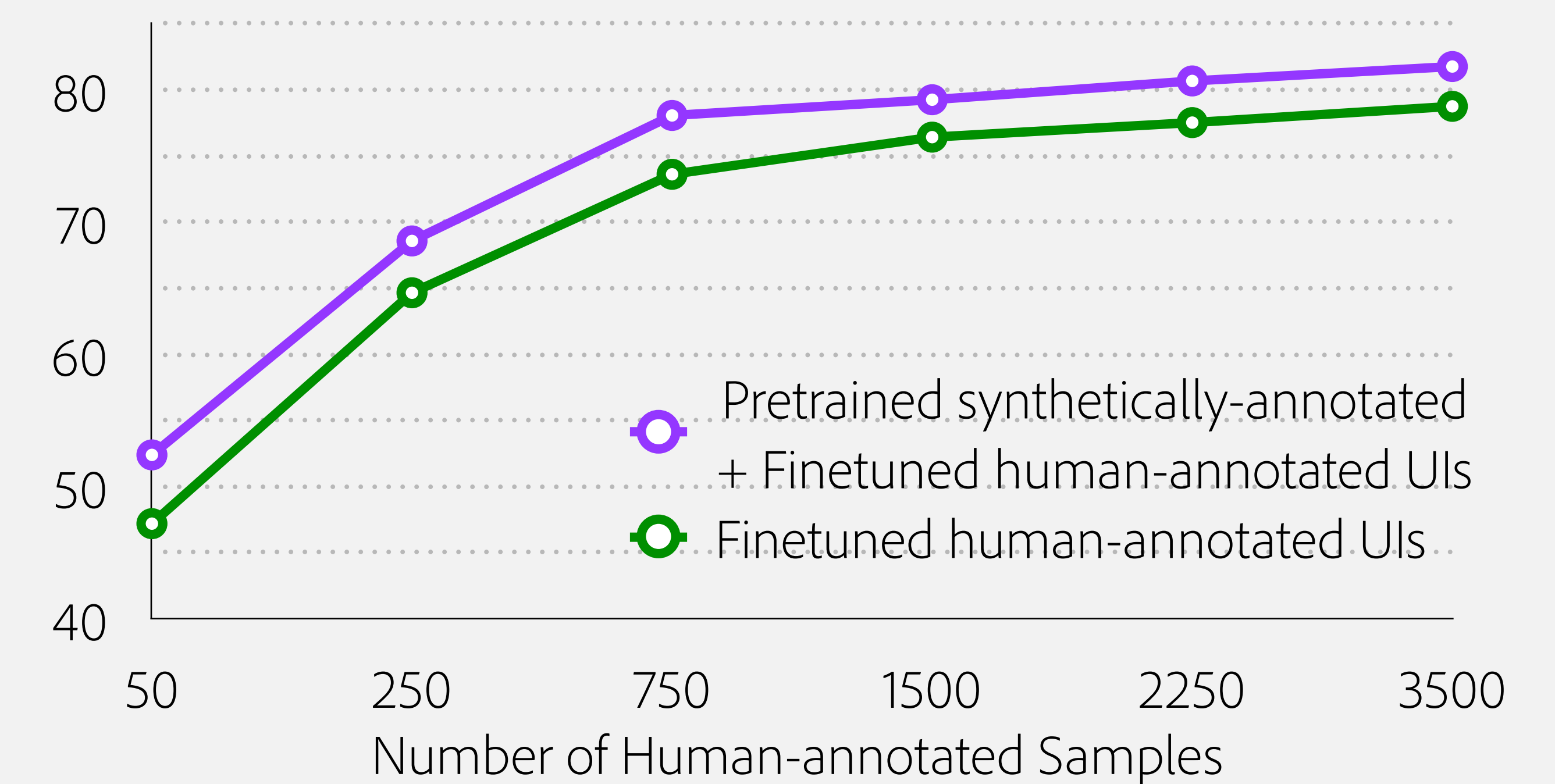


Image Caption (elo) and Classification (acc)

Model	Slides	UIs	Model	Slides	UIs
GPT-4V	1206	1235	GPT-4V	0.772	0.598
LLaVA-zero-shot	857	1123	Gemini-1.0-Pro-V	0.700	0.533
LLaVA-human-annotated	796	457	VL-zero-shot	0.467	0.388
VL-human-annotated			VL-human-annotated	0.338	0.309
LLaVA-synth (ours)	1141	1186	VL-synth (ours)	0.679	0.552