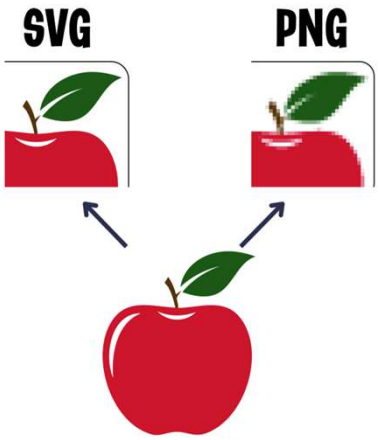
Evaluating Constrained Iterative Refinement for SVG Generation with Off-the-Shelf VLMs

Matthew Perlman

James Beetham, Dr. Amrit Bedi, Dr. Mubarak Shah

Problem Motivation

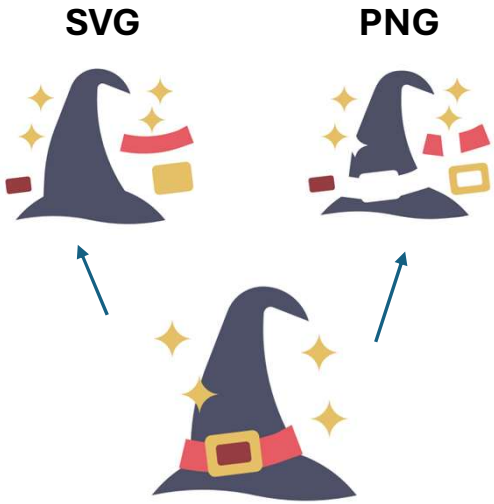
What are SVG Images? Why use off-the-shelf VLMs to generate them?



Scalable



Efficient

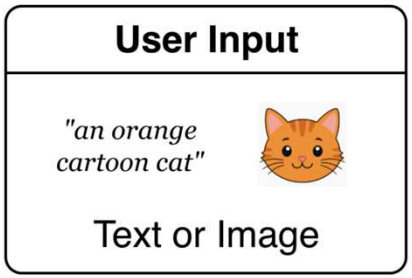


Convenient

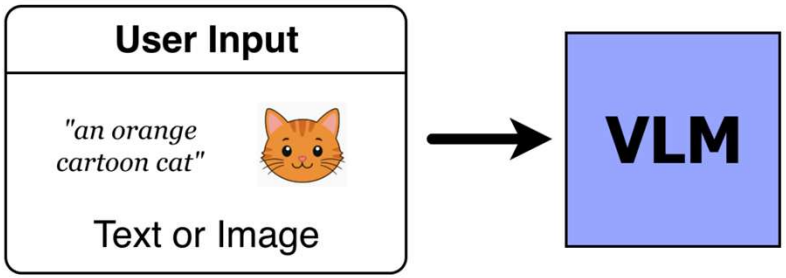
VLMs Struggle to Generate SVG Images

- Smaller models struggle to produce complex code syntax
- They produce code autoregressively, making choices irreversible
- VLMs never see the result, they guess how code will render

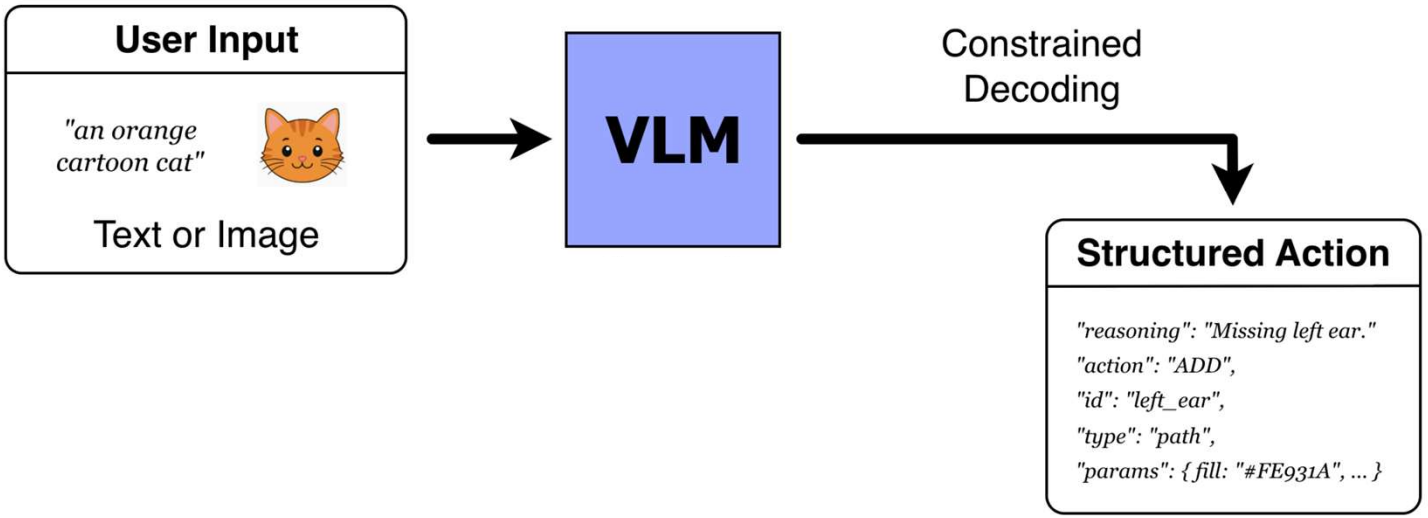
Solution - Constrained Iterative Refinement



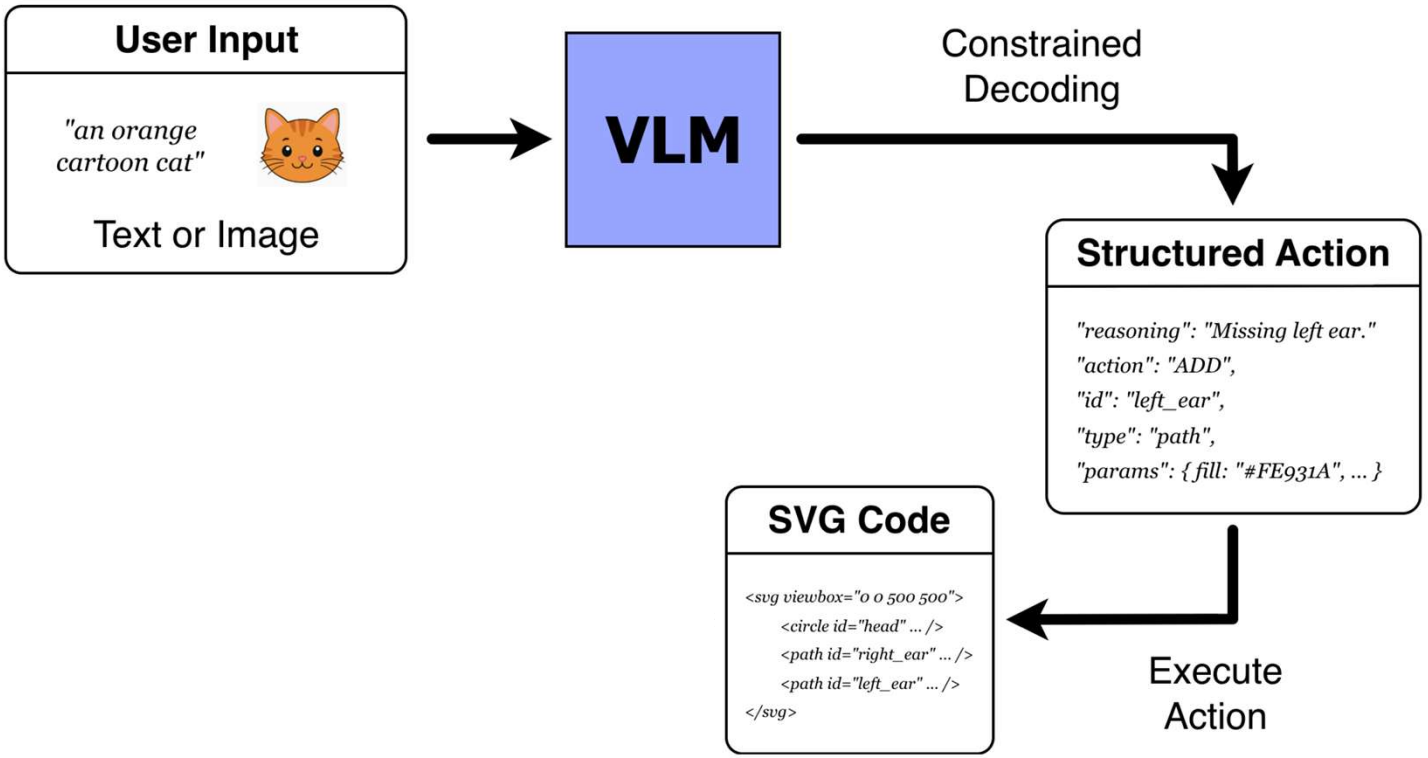
Solution - Constrained Iterative Refinement



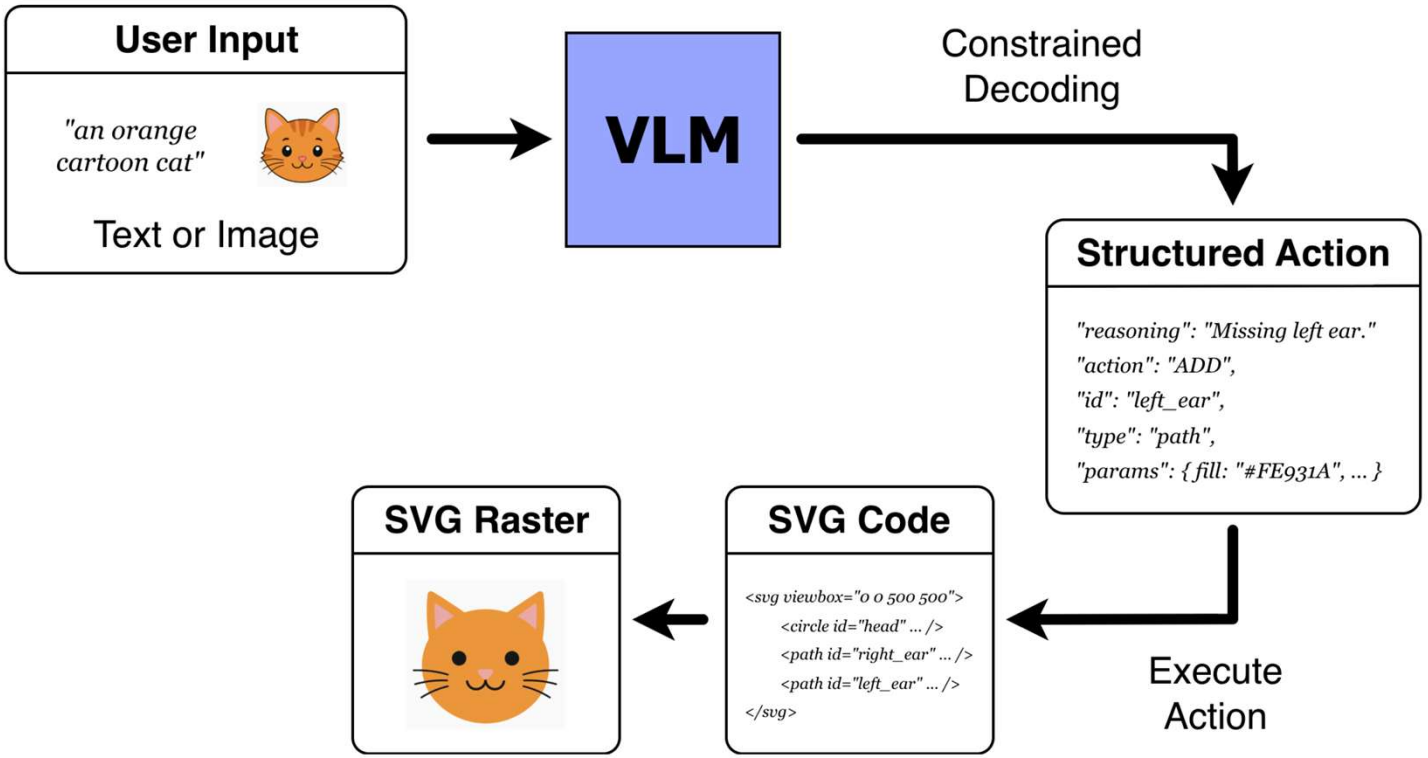
Solution - Constrained Iterative Refinement



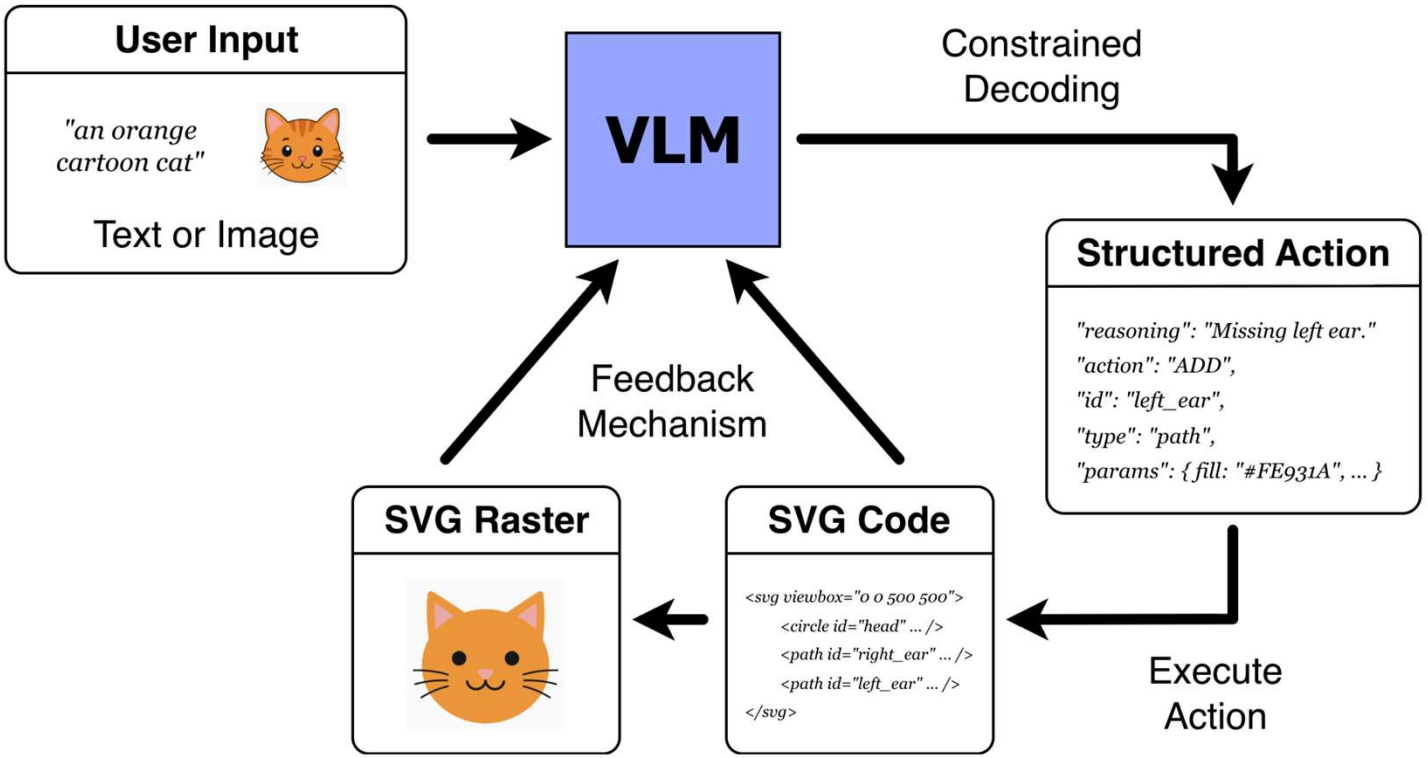
Solution - Constrained Iterative Refinement



Solution - Constrained Iterative Refinement



Solution - Constrained Iterative Refinement



Visualizing Iterative Refinement

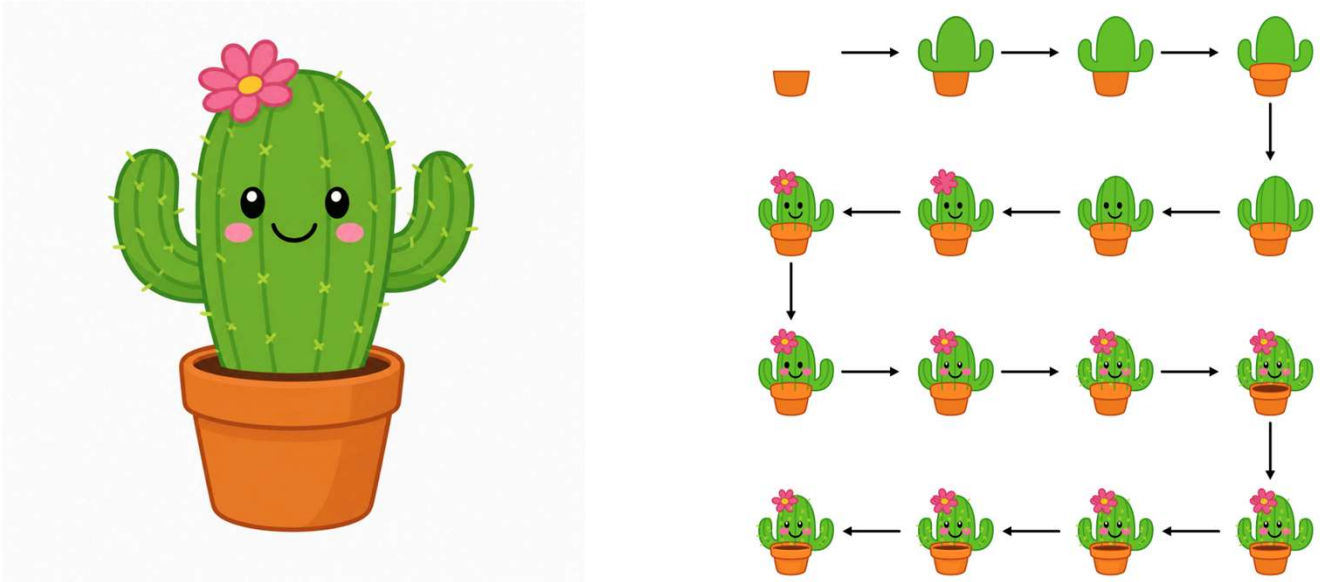
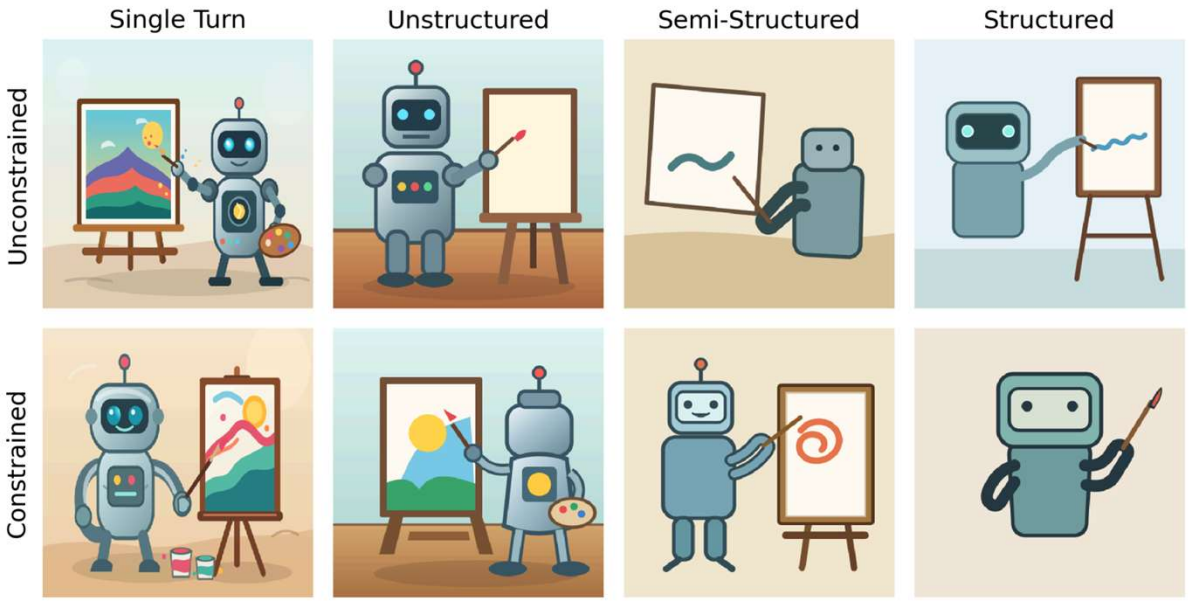


Image-to-SVG generation trajectory produced by GPT-5.6 with unconstrained structured actions. Left: reference image. Right: intermediate SVG renders across 16 iterative refinement steps.

Different Action Spaces



Images generated by GPT-5.6 using each pipeline with the Text-to-SVG prompt "A robot painting".

Metrics on Qwen3.5

Constrained Decoding Consistently Improves Compilation.

Image Quality and Prompt Adherence are competitive with SOTA.

		Text→SVG				Image→SVG			
Decoding	**Pipeline**	**Compile↑**	**CLIP↑**	**Aesthetic.↑**	**#Tokens**	**Compile↑**	**DINO↑**	**LPIPS↓**	**#Tokens**
Unconstrained	Single Turn	**0.837**	**0.255**	**4.655**	1.4k	0.630	**0.704**	0.396	1.7k
	Unstructured	0.475	0.246	4.560	4.8k	0.113	**0.704**	**0.372**	10.7k
	Semi-Structured	0.180	0.205	4.415	7.9k	0.500	0.617	0.418	8.8k
	Structured	0.785	0.226	4.480	3.3k	**0.650**	0.633	0.413	8.1k
Constrained	Single Turn	**0.830**	0.220	**4.530**	0.6k	**0.830**	0.411	0.579	0.4k
	Unstructured	0.599	**0.242**	4.485	5.2k	0.647	**0.617**	0.430	5.1k
	Semi-Structured	0.512	0.206	4.445	2.7k	0.375	0.464	0.506	3.2k
	Structured	0.790	0.218	4.440	2.4k	0.695	0.603	**0.420**	3.3k

Results averaged across Qwen3.5 9B and 27B on MMSVG-Bench.