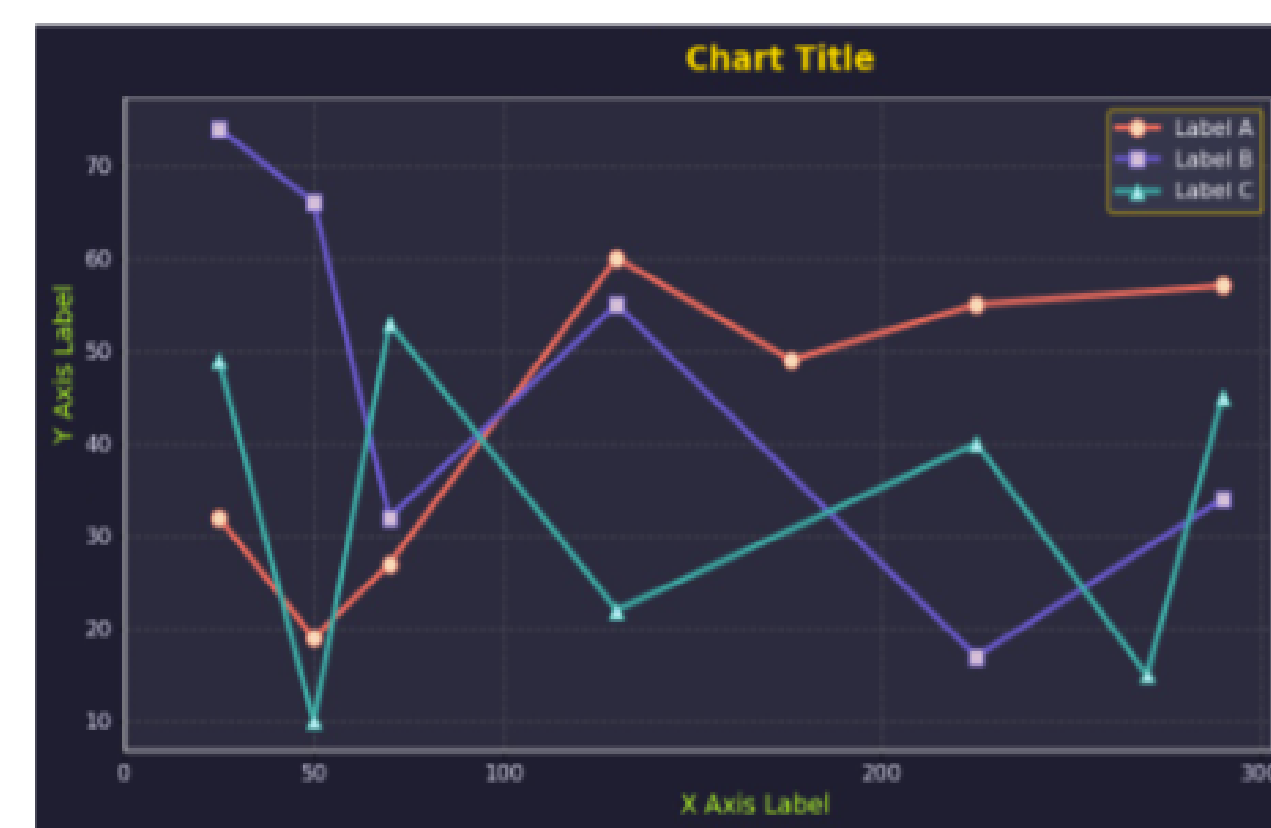


Background

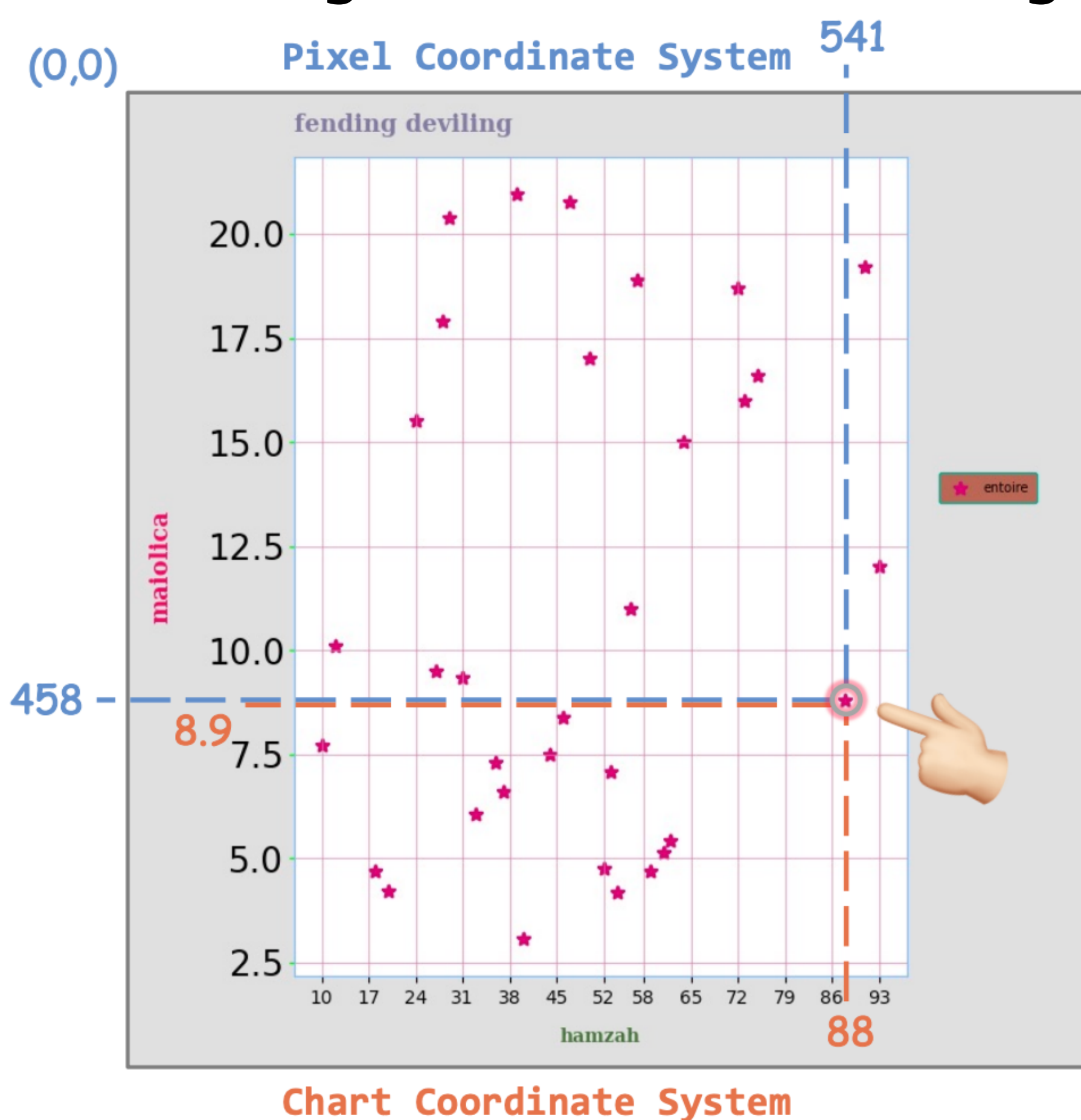
- Existing models face several issues, such as large numerical errors, misaligned data, missing points, or hallucinations.
- Current Reasoning LLMs mostly rely on raw images and plain text for reasoning.
- For tasks where the difficulty is "visual", text-based reasoning is often ineffective.



"Label A"... "Label B"(25,74)(50,66)(70,32)(130,55)(225,17)(290,34) "Label C"...
 "Label A"... "Label B"(0,70)(50,60)(100,50)(150,40)(200,50)(250,60)(300,70) "Label C"...
 "Label A"(25,32)(50,19)(70,27)(130,60)(176,49)(225,55)(290,57) "Label B"... "Label C"...
 "Label A"(50,30)(60,20)(75,25)(125,60) **missing** (225,55)(275,55) "Label B"... "Label C"...

Motivation

- Humans typically pinpoint a data point's location and then read its coordinates.
- Visualizing intermediate reasoning steps as "visual thought" is more effective.

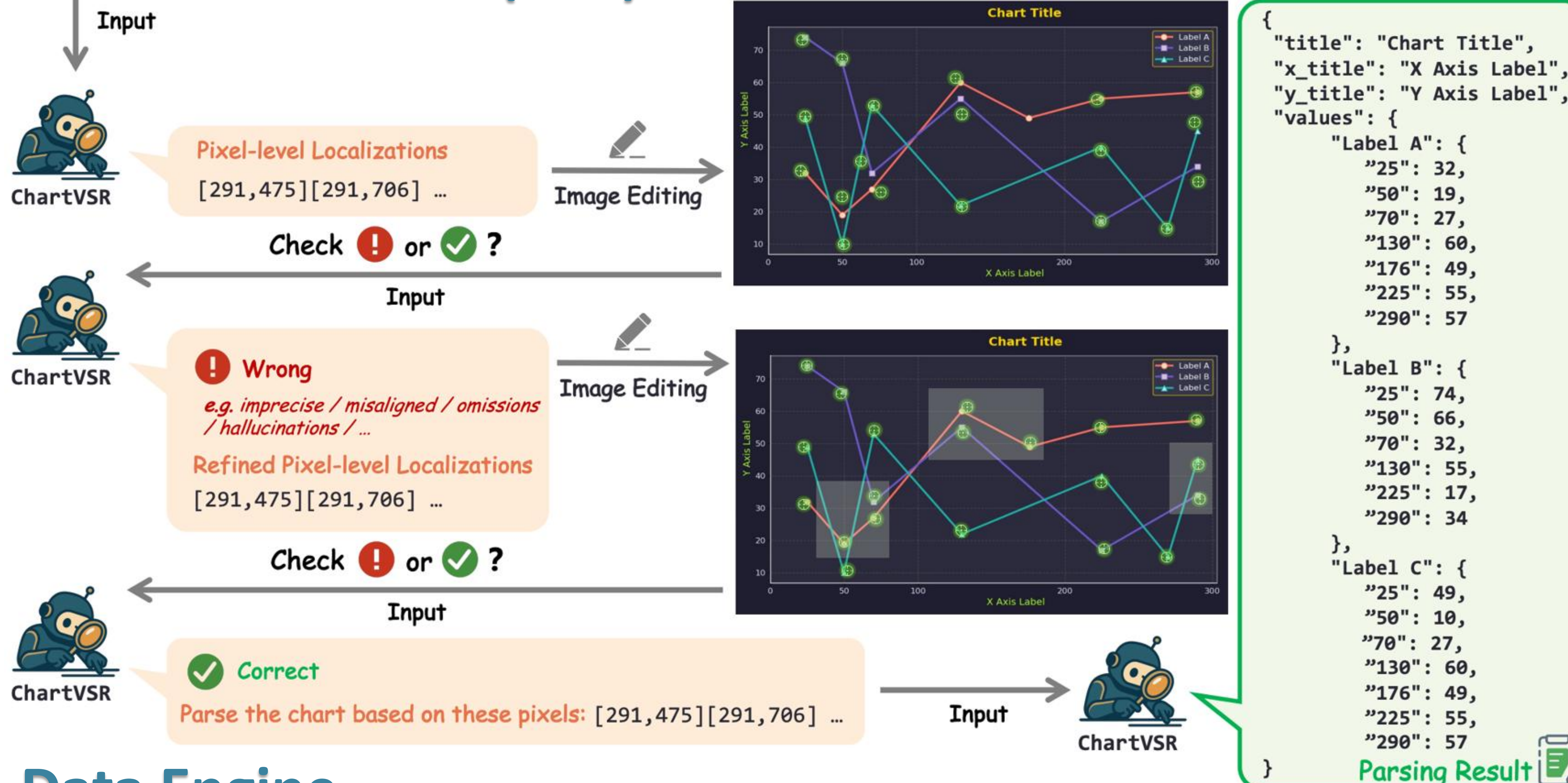


- Pixel-level Localization: [541, 458]
- Chart Coordinates: (88.0, 8.9)

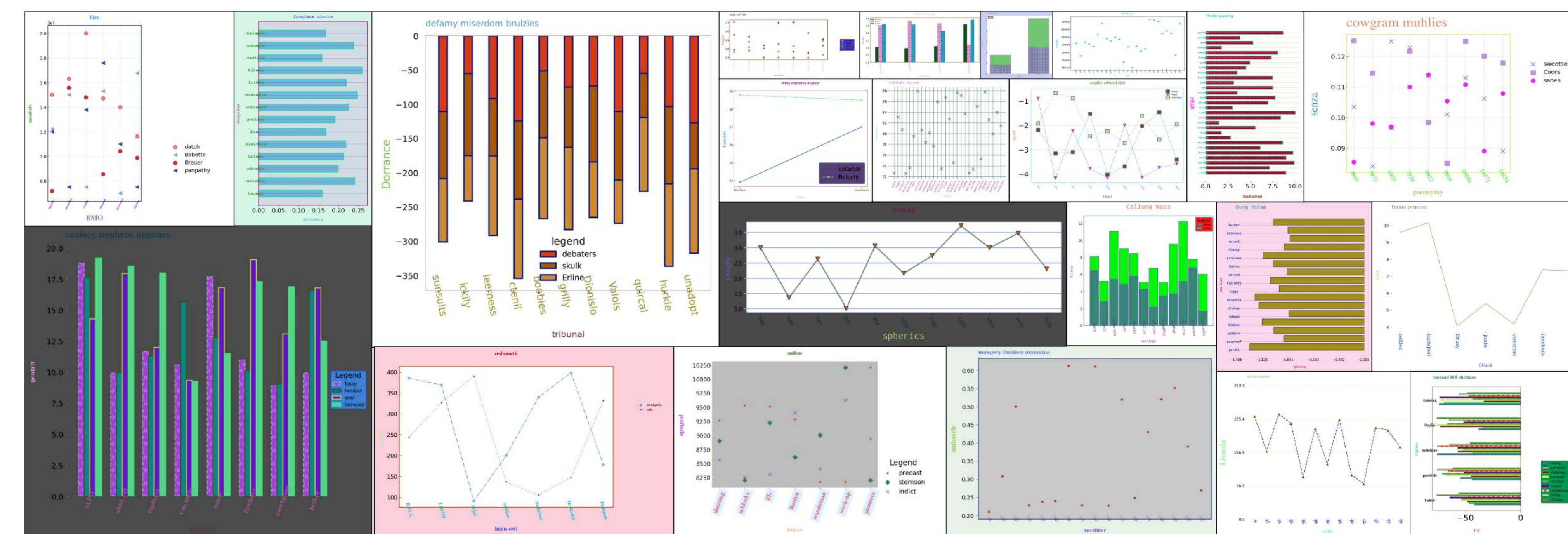
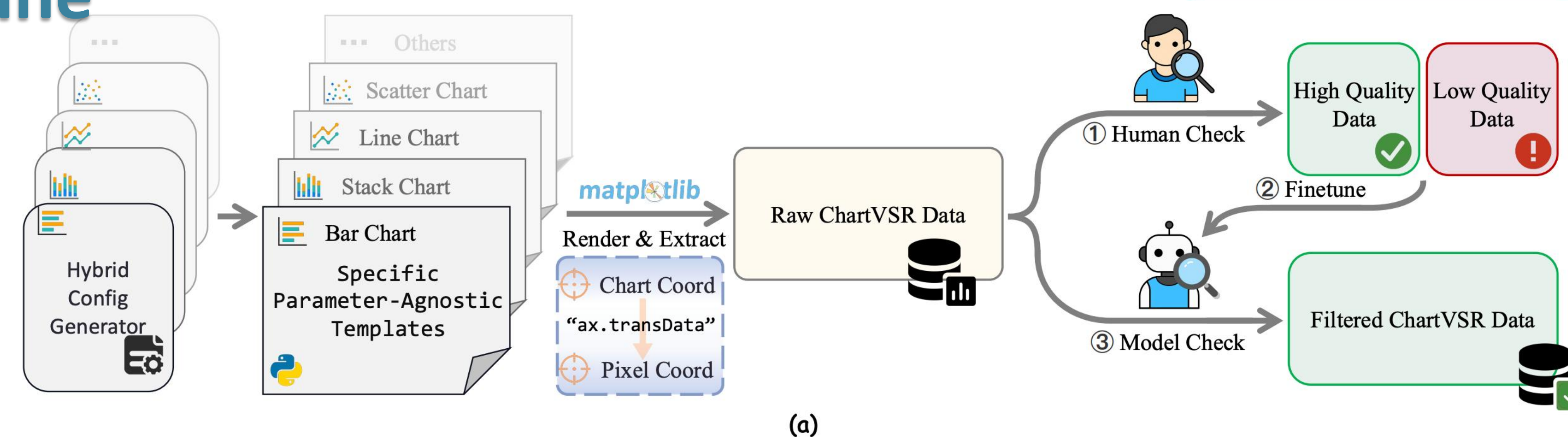
Model	Size	ChartQA-SE-Clean			PlotQA-SE			ChartX-SE		
		AP-Strict	AP-Slight	AP-High	AP-Strict	AP-Slight	AP-High	AP-Strict	AP-Slight	AP-High
DePlot	1.3B	55.61	66.89	70.89	3.11	16.49	26.50	14.84	33.42	42.42
ChartAst	13B	33.81	63.82	71.23	5.18	48.67	56.08	18.78	36.76	47.28
ChartVLM	7.3B	65.94	77.17	82.11	3.81	46.83	54.00	25.42	34.92	40.82
OneChart	0.2B	66.20	78.89	83.92	34.56	84.18	86.10	44.55	53.12	59.72
ChartVSR	3B	65.70	83.69	85.64	34.99	84.61	88.10	43.99	53.48	62.89

Model	Size	Easy Subset			Avg.	Hard Subset			Avg.	
		AP-Strict	AP-Slight	AP-High		AP-Strict	AP-Slight	AP-High		
<i>Strong Closed-source LLMs</i>										
GPT-4o	-	0.00	1.69	6.73	2.81	0.00	0.99	4.14	1.71	2.09
Gemini-2.5-Flash	-	0.00	37.00	55.57	30.86	0.00	23.84	40.22	21.35	24.62
Gemini-2.5-Pro	-	0.06	36.51	46.05	27.54	0.00	46.11	66.39	37.50	34.07
<i>Chart LLMs</i>										
DePlot	1.3B	0.00	9.73	15.35	8.26	0.00	1.27	4.19	1.82	4.04
ChartAst	13B	0.00	1.56	5.79	2.45	0.00	0.13	3.10	1.08	1.55
ChartVLM	7.3B	0.00	1.74	6.34	2.70	0.00	0.48	3.56	1.35	1.81
OneChart	0.2B	0.00	4.75	18.89	7.88	0.00	1.98	9.48	3.82	5.22
ChartVSR	3B	0.11	51.95	67.44	39.83	0.02	49.30	63.67	37.66	38.41

Visual Self-Refine (VSR)



Data Engine



Presenter: Jinsong Li

Feel free to reach out ~

