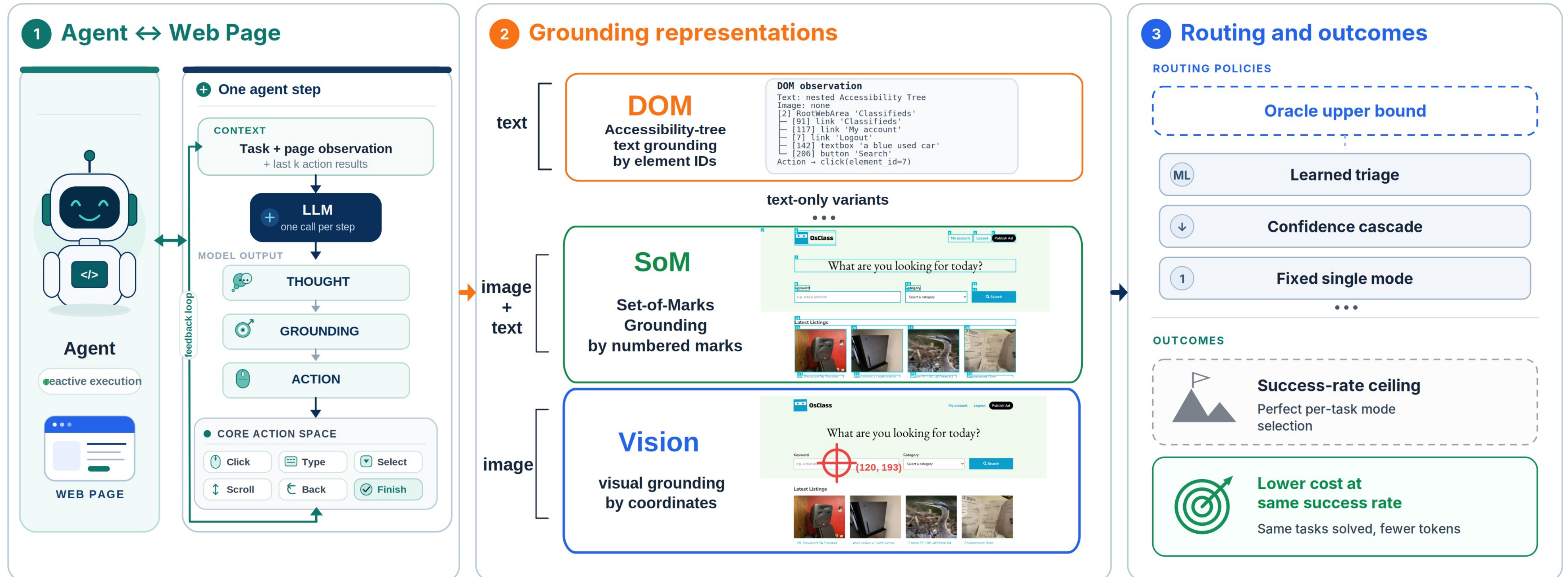


Testing when richer web-agent representations help — and whether their value can be predicted cheaply

WHAT AN AGENT SEES, AND WHO CHOOSES



THE SAME TASK, SEEN TWO WAYS

“Go to my listing of the blue bike and change the price to \$85.50 — and say so in the description.”

READ
page text only
12 steps
8 pages seen
\$0.09 solved

LOOK
screenshot only
26 steps
9 pages seen
\$0.10 gave up

Same start. READ finishes; LOOK loops back. Both outcomes repeat on an independent rerun.

step 4 · type → \$0.020 spent

step 8 · type → \$0.050 spent

step 10 · click → \$0.086 spent

step 2 · click → \$0.011 spent

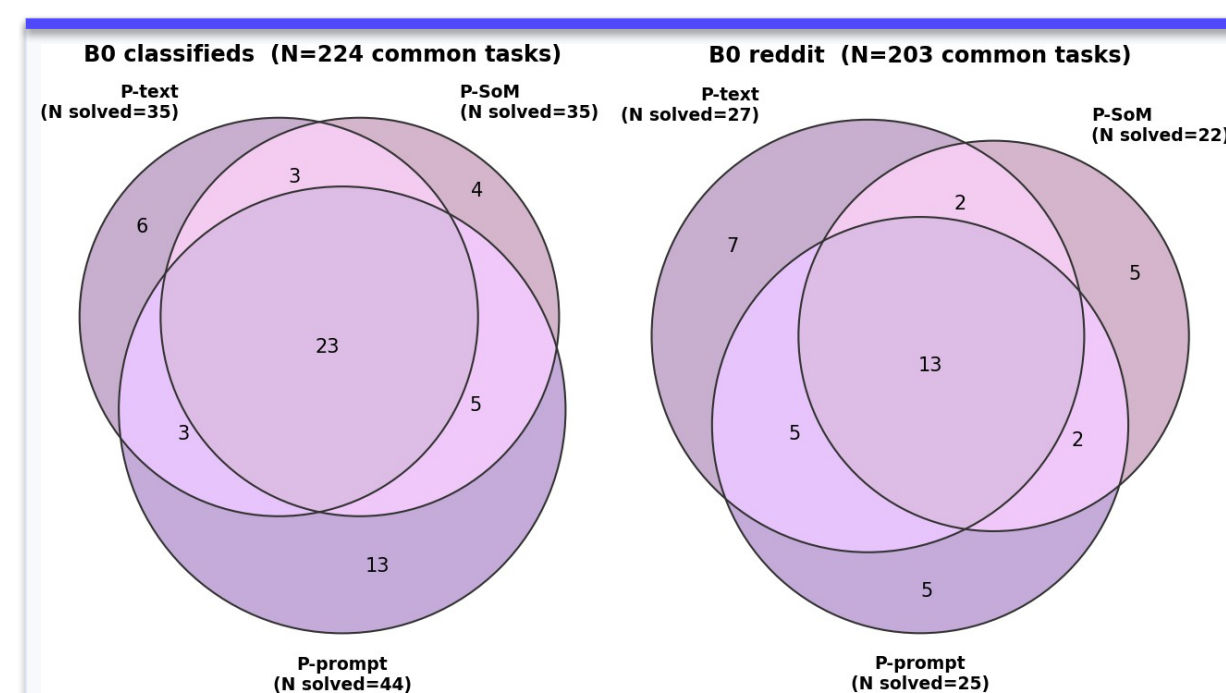
step 17 · click → \$0.071 · left the form

step 19 · click → \$0.079 · back on frame 1

1 · NO VIEW WINS EVERYWHERE

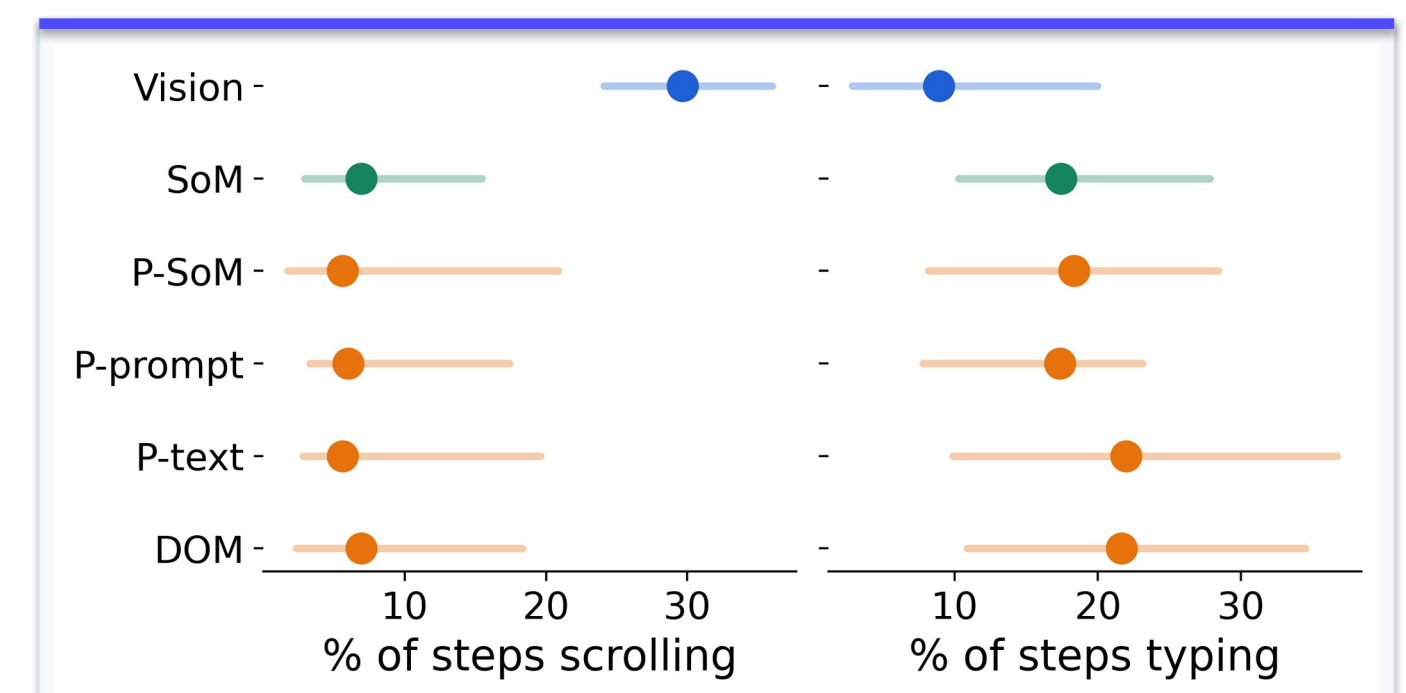
	VisualWebArena								WebArena	
	DOM	P-prompt	P-text	P-SoM	SoM	Vision	no image	text + image	image only	
cls-80 n=224	17.4	14.3	6.2	5.9	1.3	3.9	26.9	16.4		
cls-80 n=203	19.4	12.3	6.7	5.4	1.8	0.0	26.0	16.4		
cls-81 n=224	15.6	13.3	7.6	5.9	0.5	2.0	35.6	16.4		
cls-81 n=203	15.6	10.8	6.7	5.9	0.9	0.5	25.0	11.5		
cls-82 n=224	27.2	14.8	14.3	7.4	2.2	1.0	22.1	13.5		
cls-82 n=203	25.0	7.4	12.5	2.5	2.2	2.0	19.2	9.6		
cls-83 n=104										
cls-83 n=104										

2 · THEY SOLVE DIFFERENT TASKS



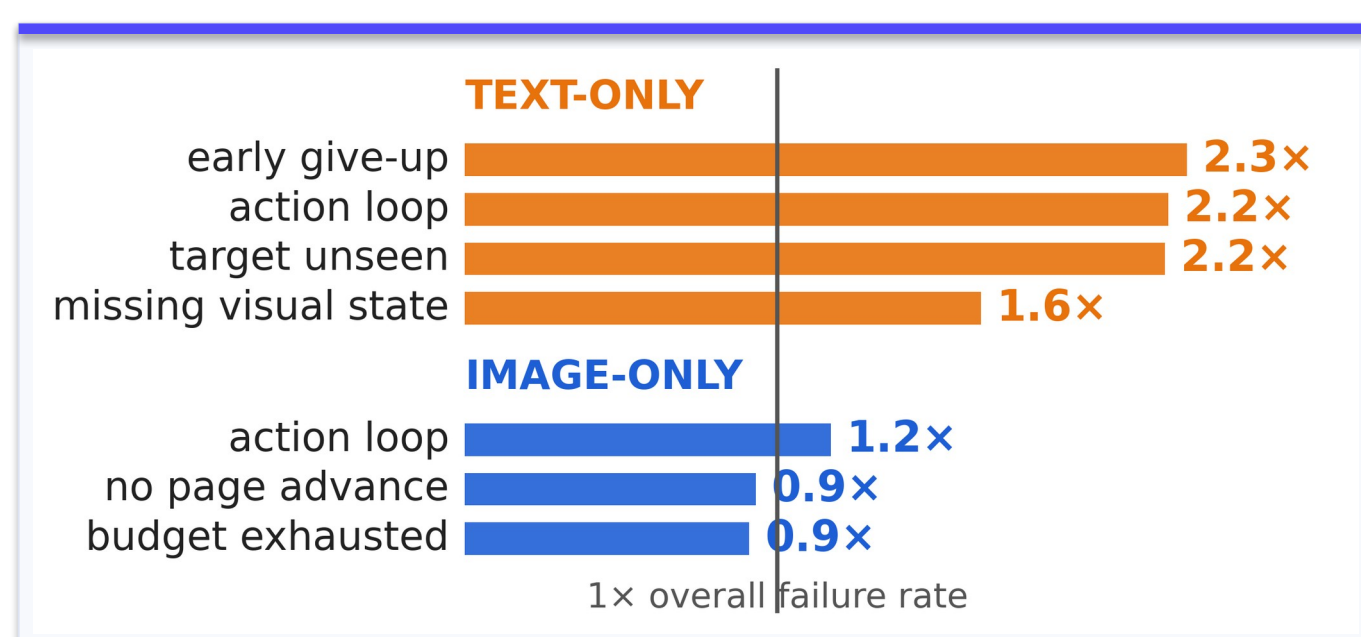
The sets overlap — but do not coincide.

3 · THEY BEHAVE DIFFERENTLY



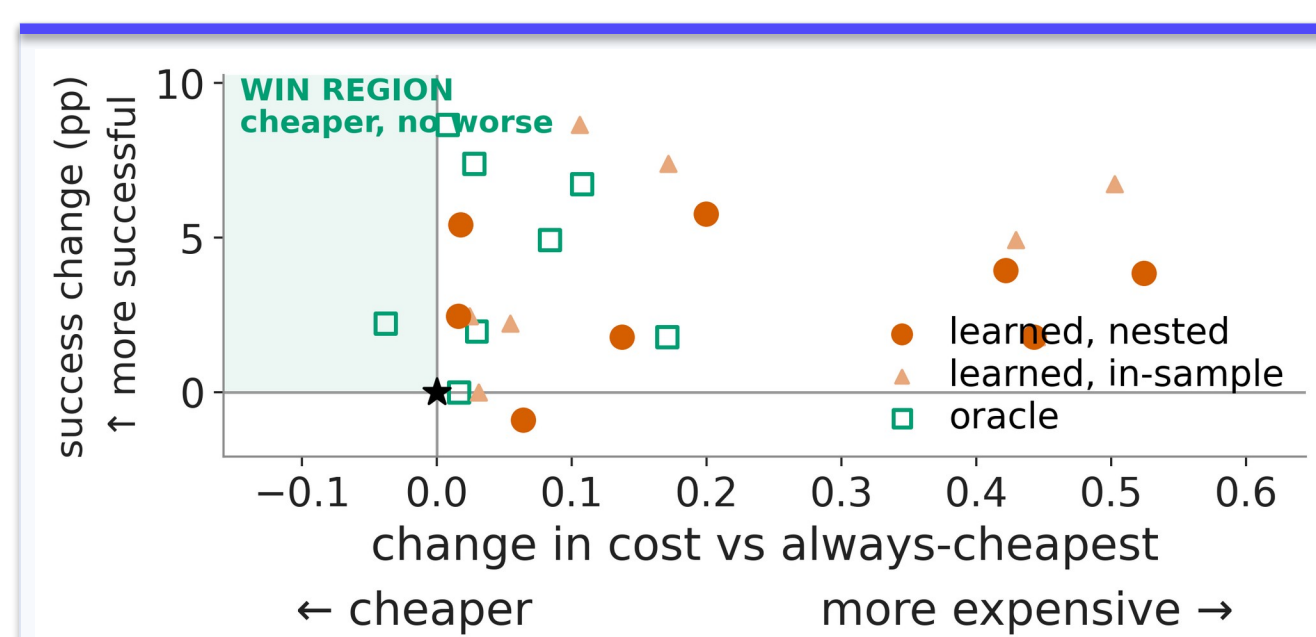
Vision scrolls ~4x more.

4 · AND THEY FAIL DIFFERENTLY



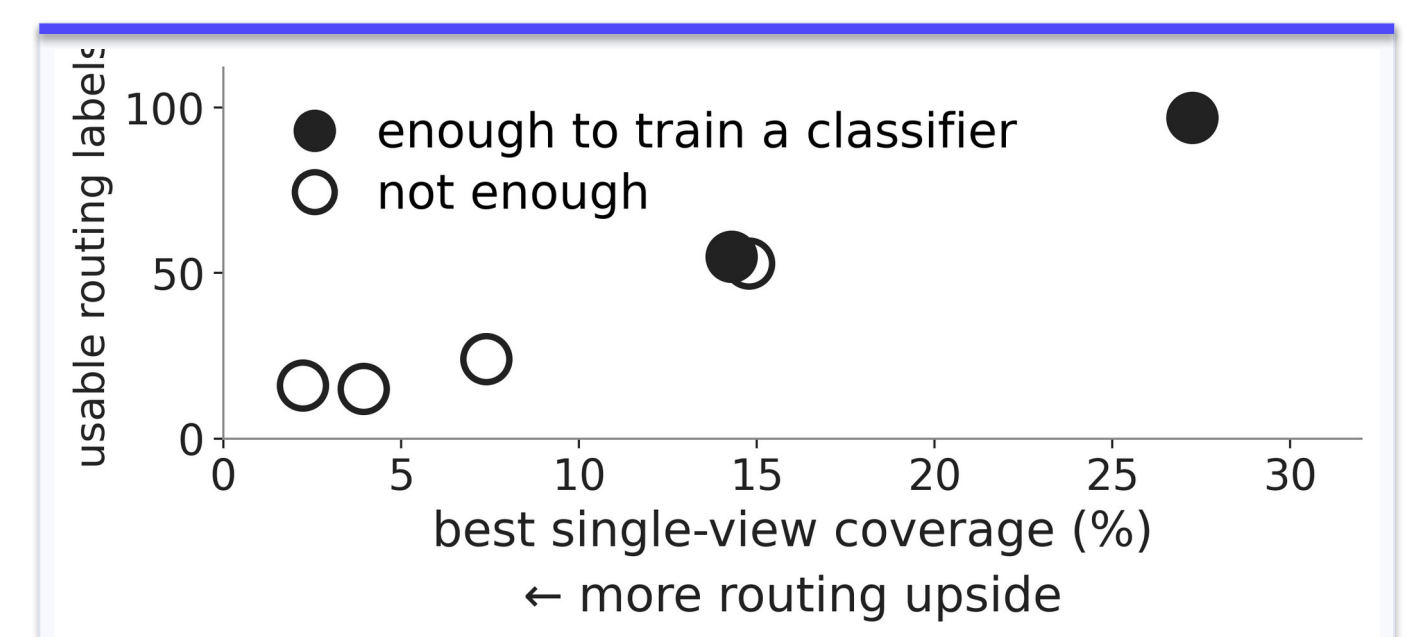
Text-only skews toward early give-up; image-only toward stalled progress.

5 · SO BUILD A ROUTER? NOT SO FAST.



Learned routers buy success only by spending more; only the hindsight oracle reaches the win region.

6 · AND THIS IS WHY



More routing upside, less usable training signal.

