Average Steps Comparison

Legend:
- Baseline
- WebCoach (GPT-4o)
- WebCoach (Qwen3-8b)
- WebCoach (Qwen3) with Self-exp
- GPT-4o baseline

Y-axis: Average Steps

X-axis: Base Model

qwen_vl_7b: 16.4, 16.6, 16.4, 17.4

qwen_vl_32b: 13.3, 10.9, 12.2, 11.9

skywork_r1v3: 10.7, 10.7, GPT-4o 10.9, 10.3, 10.2