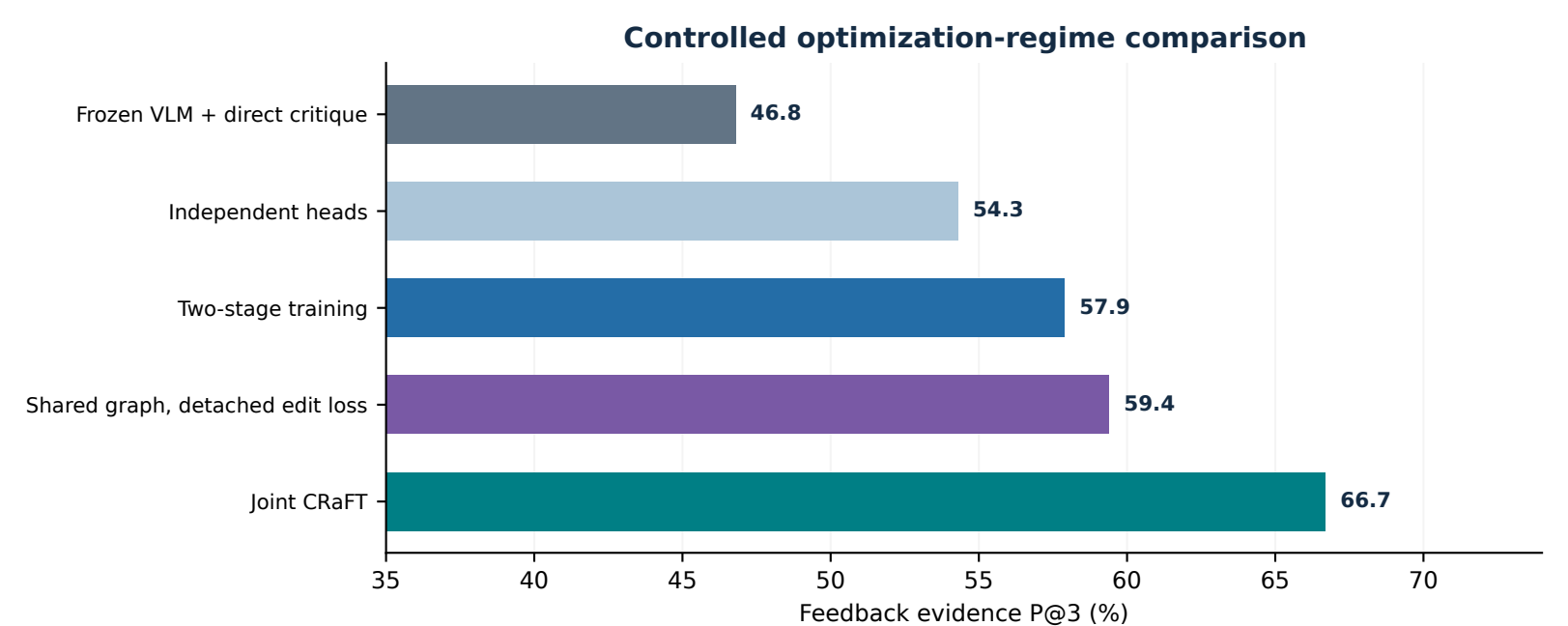

Controlled optimization-regime comparison
Frozen VLM + direct critique
46.8
Independent heads
54.3
Two-stage training
57.9
Shared graph, detached edit loss
59.4
Joint CRaFT
66.7
35
40
45
50
55
60
65
70
Feedback evidence P@3 (%)