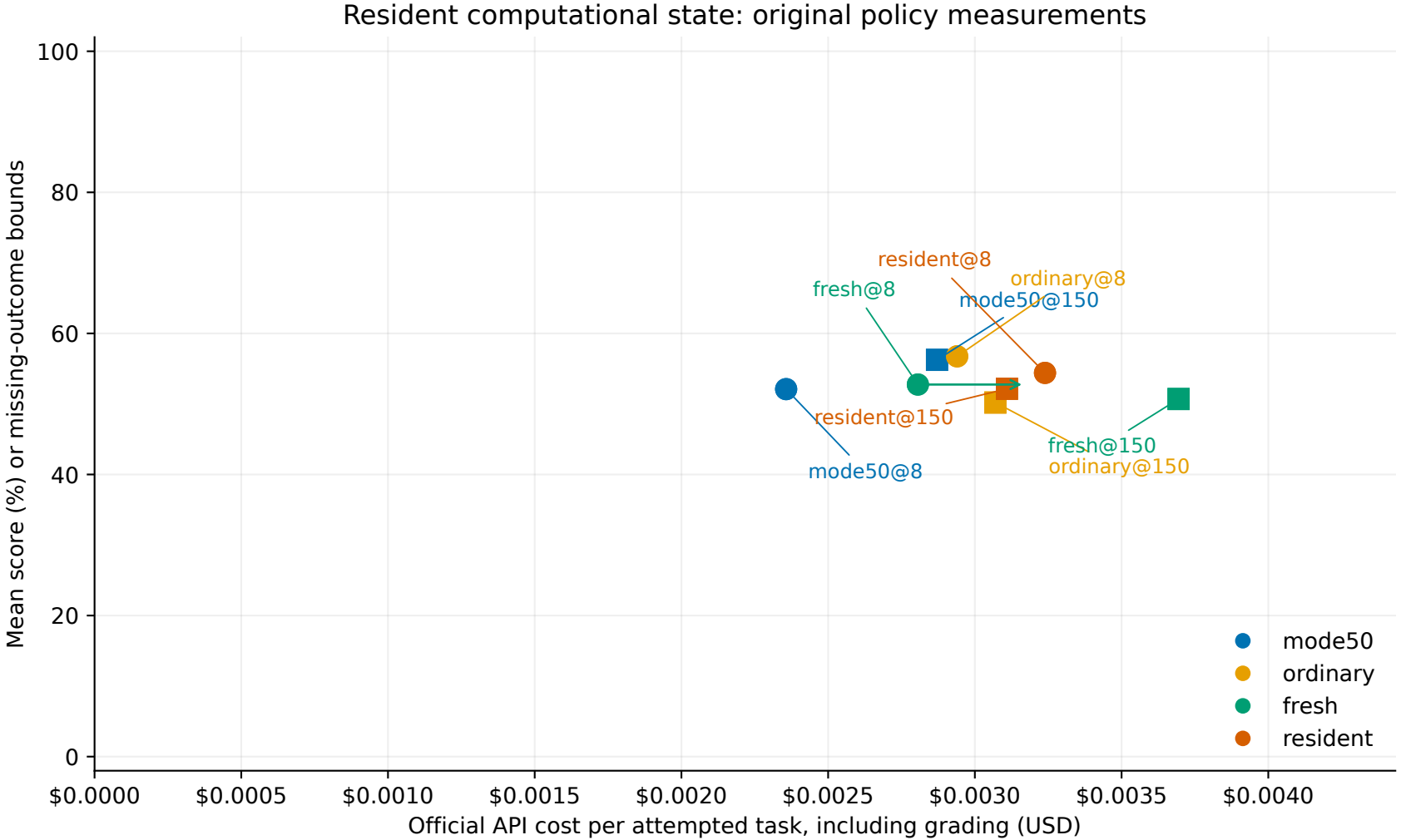

Resident computational state: original policy measurements
Mean score (%) or missing-outcome bounds
100
80
60
40
20
0
$0.0000
$0.0005
$0.0010
$0.0015
$0.0020
$0.0025
$0.0030
$0.0035
$0.0040
Official API cost per attempted task, including grading (USD)
resident@8
ordinary@8
fresh@8
mode50@150
resident@150
fresh@150
ordinary@150
mode50@8
mode50
ordinary
fresh
resident
Vertical spans: missing-outcome bounds. Right arrows: cost lower bounds. ×: protocol errors.
Points are not interpolated; no confidence bands or population frontier are implied.