deepseek-r1-70b - A1

llama3 - A1

deepseek-r1-70b - A2

llama3 - A2

deepseek-r1-70b - A2_older

llama3 - A2_older

deepseek-r1-70b - B1

llama3 - B1

deepseek-r1-70b - B2

llama3 - B2

deepseek-r1-70b

llama3

precision: 0.11 0.81
recall: 0.04 0.92
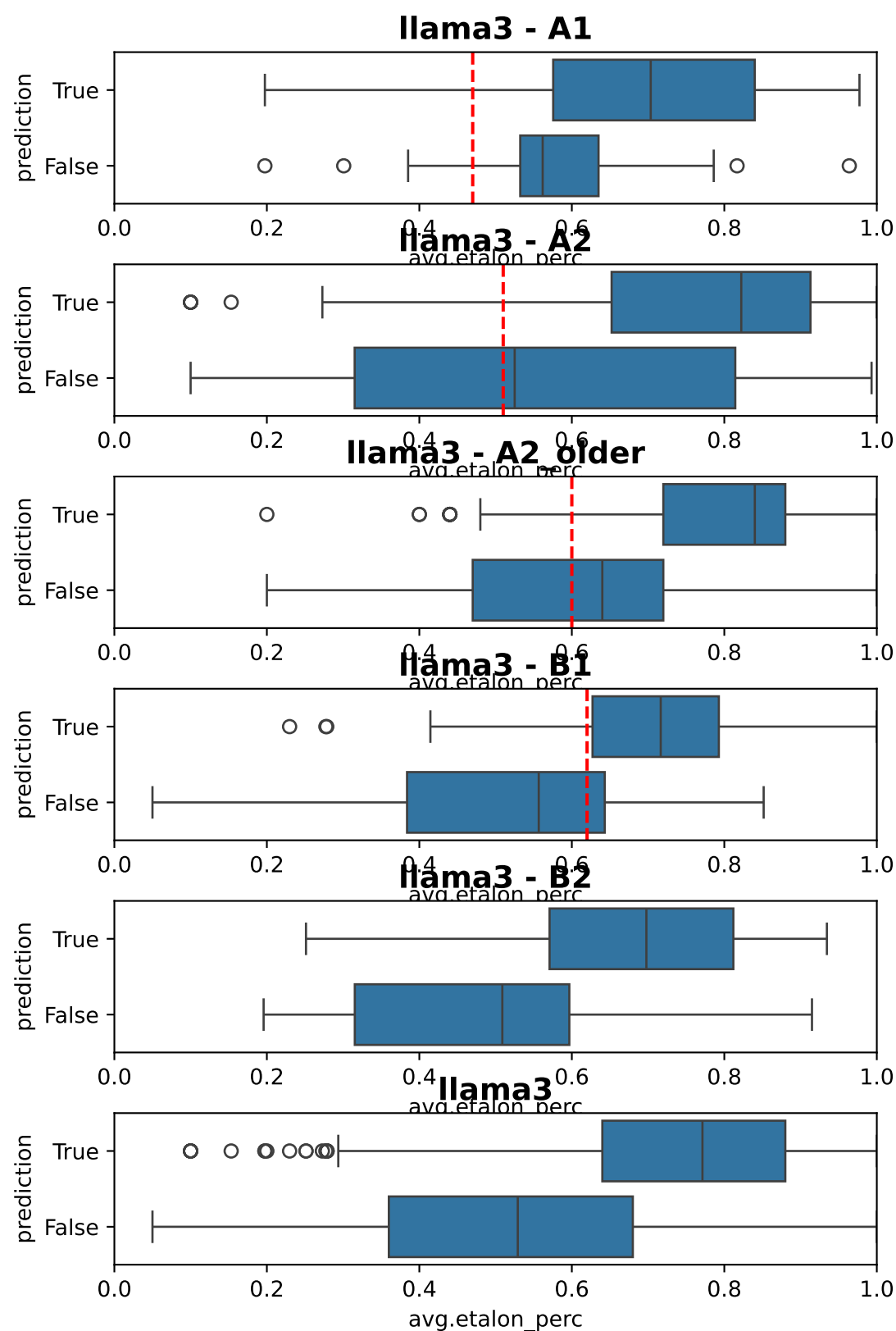f1-score: 0.06 0.86
support: 24 105
QWK: -0.05

precision: 0.24 0.83
recall: 0.25 0.82
f1-score: 0.24 0.82
support: 24 105
QWK: 0.07

precision: 0.38 0.78
recall: 0.12 0.94
f1-score: 0.18 0.85
support: 90 294
QWK: 0.08

precision: 0.46 0.87
recall: 0.61 0.78
f1-score: 0.53 0.82
support: 90 294
QWK: 0.35

precision: 0.33 0.85
recall: 0.06 0.98
f1-score: 0.10 0.91
support: 36 204
QWK: 0.05

precision: 0.42 0.90
recall: 0.42 0.90
f1-score: 0.42 0.90
support: 36 204
QWK: 0.31

precision: 1.00 0.62
recall: 0.02 1.00
f1-score: 0.04 0.77
support: 111 180
QWK: 0.02

precision: 0.65 0.75
recall: 0.56 0.81
f1-score: 0.60 0.78
support: 111 180
QWK: 0.38

precision: 0.60 0.70
recall: 0.12 0.96
f1-score: 0.20 0.81
support: 51 108
QWK: 0.10

precision: 0.49 0.80
recall: 0.63 0.69
f1-score: 0.55 0.74
support: 51 108
QWK: 0.30

precision: 0.38 0.81
recall: 0.09 0.96
f1-score: 0.14 0.88
support: 234 969
QWK: 0.08

precision: 0.42 0.89
recall: 0.61 0.79
f1-score: 0.49 0.84
support: 234 969
QWK: 0.34