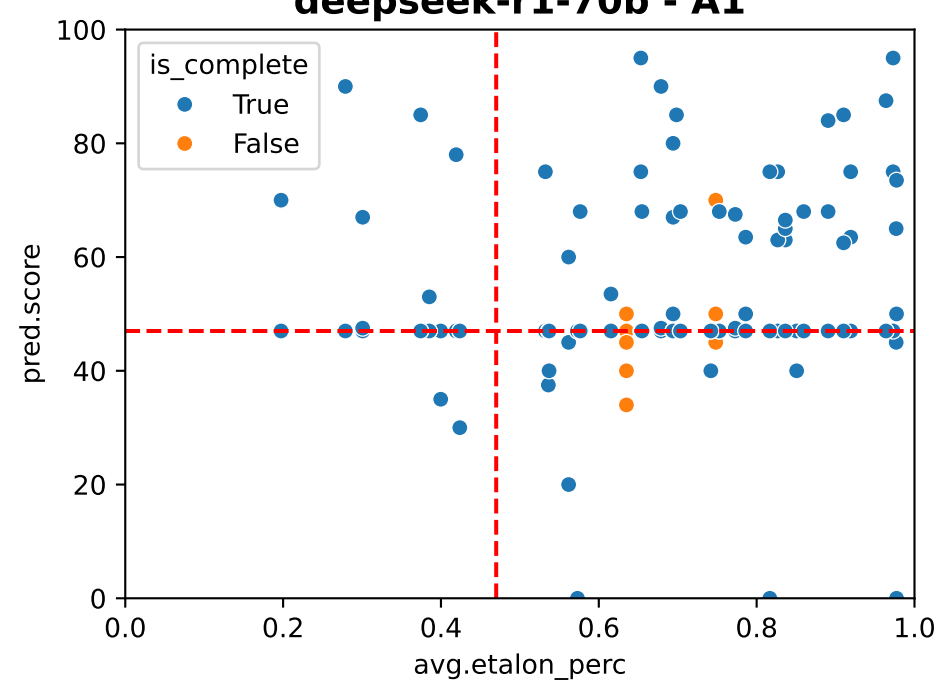
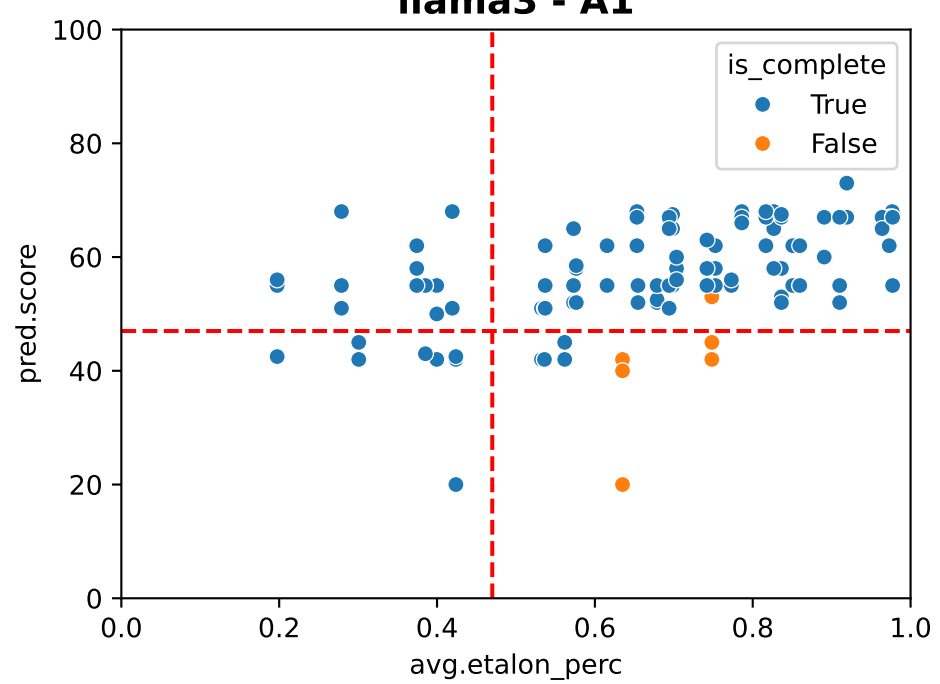


**deepseek-r1-70b - A1**



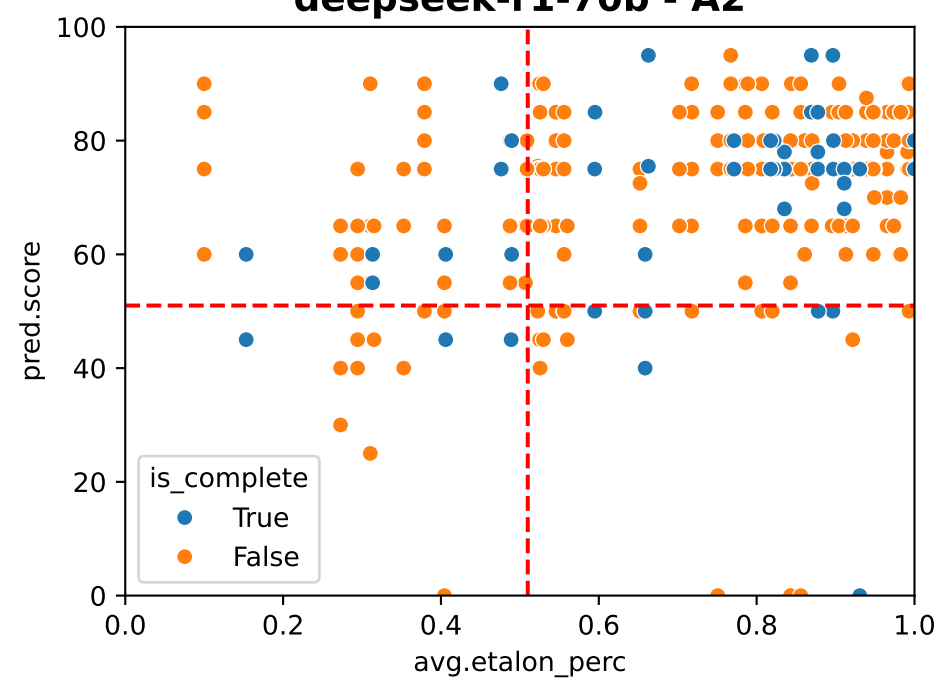
precision: 0.12 0.80  
recall: 0.08 0.86  
f1-score: 0.10 0.83  
support: 24 105  
QWK: -0.07

**llama3 - A1**



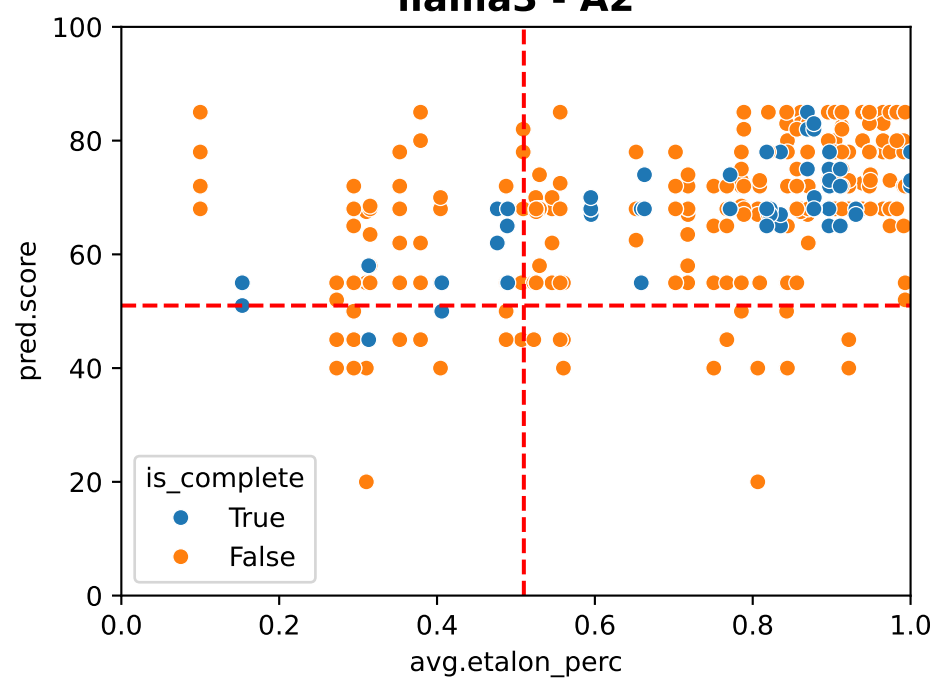
precision: 0.36 0.86  
recall: 0.38 0.85  
f1-score: 0.37 0.85  
support: 24 105  
QWK: 0.22

**deepseek-r1-70b - A2**



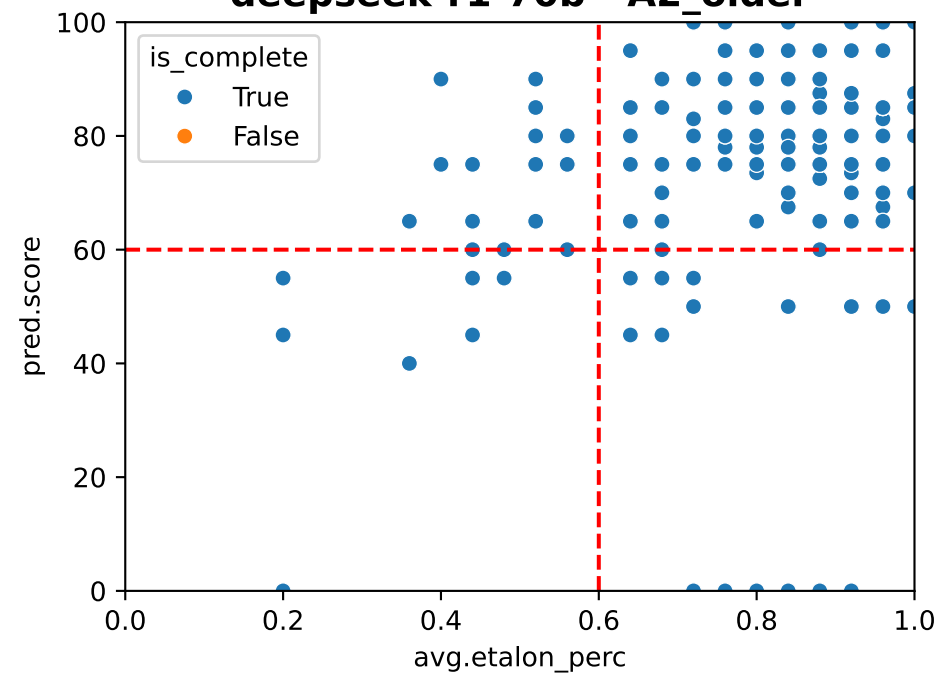
precision: 0.35 0.78  
recall: 0.17 0.90  
f1-score: 0.23 0.84  
support: 90 294  
QWK: 0.09

**llama3 - A2**



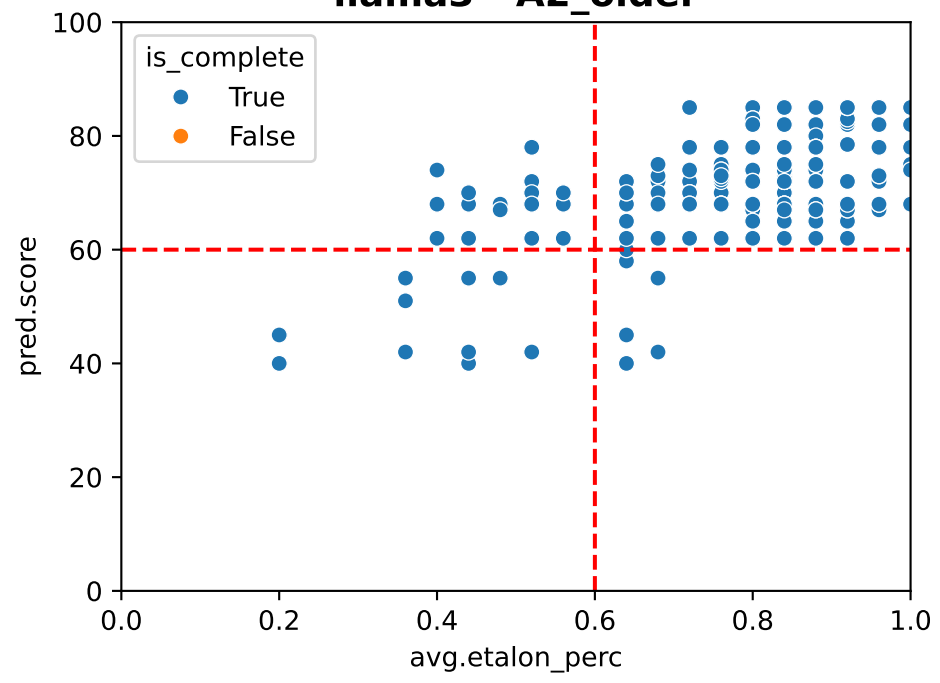
precision: 0.53 0.80  
recall: 0.26 0.93  
f1-score: 0.35 0.86  
support: 90 294  
QWK: 0.23

**deepseek-r1-70b - A2\_older**



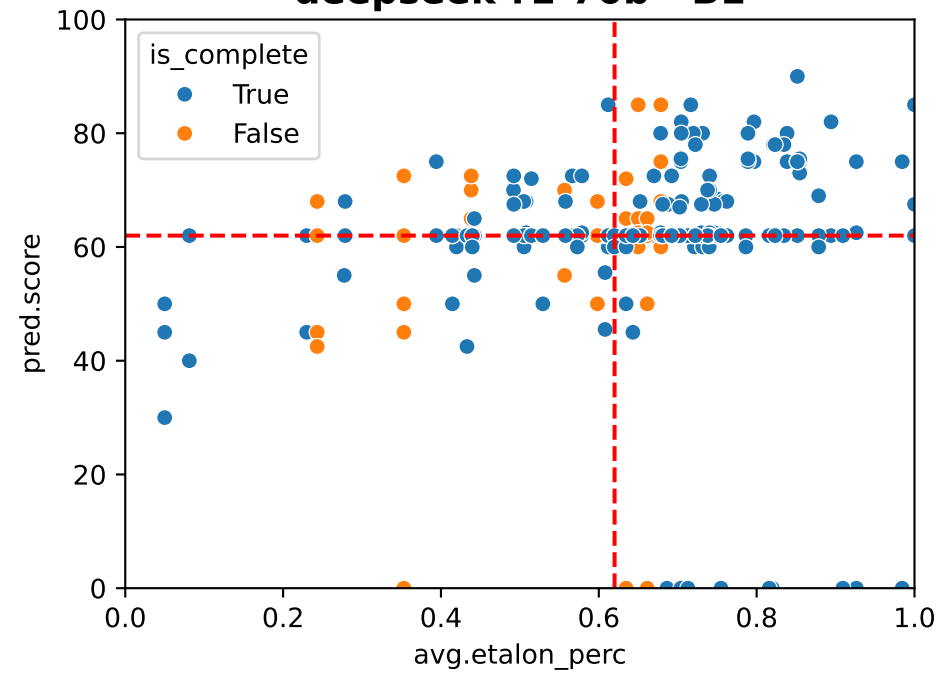
precision: 0.26 0.87  
recall: 0.22 0.89  
f1-score: 0.24 0.88  
support: 36 204  
QWK: 0.12

**llama3 - A2\_older**



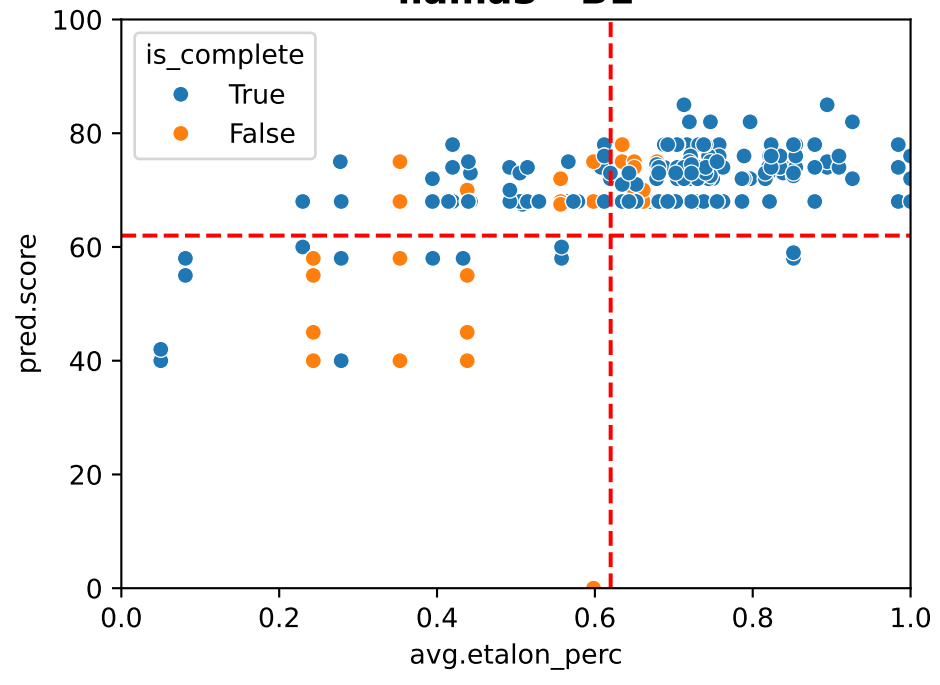
precision: 0.65 0.90  
recall: 0.36 0.97  
f1-score: 0.46 0.93  
support: 36 204  
QWK: 0.40

**deepseek-r1-70b - B1**



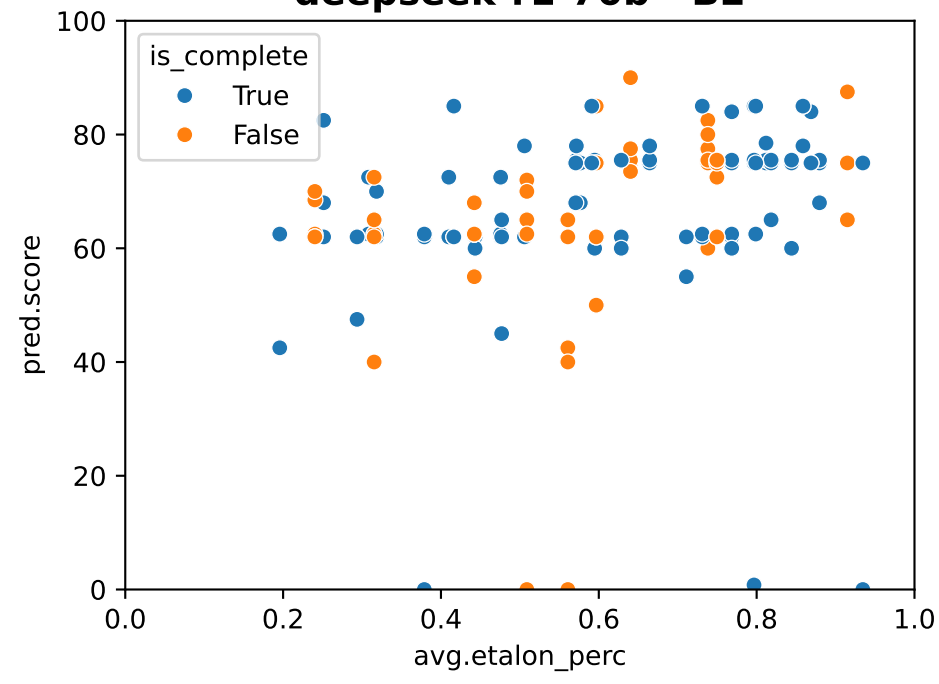
precision: 0.53 0.65  
recall: 0.23 0.88  
f1-score: 0.32 0.75  
support: 111 180  
QWK: 0.12

**llama3 - B1**



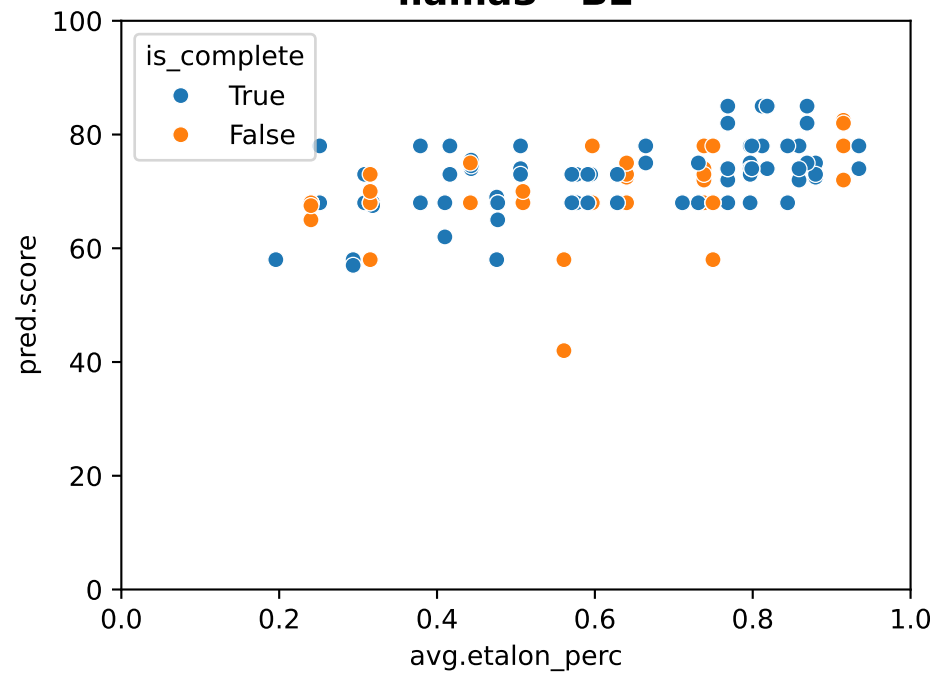
precision: 0.92 0.70  
recall: 0.31 0.98  
f1-score: 0.46 0.82  
support: 111 180  
QWK: 0.33

**deepseek-r1-70b - B2**



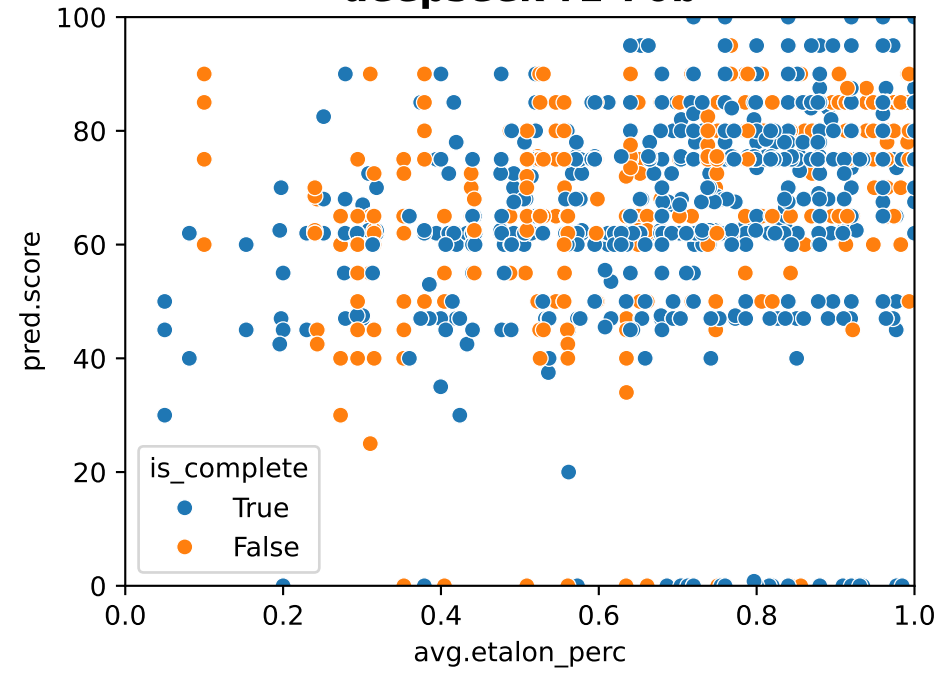
precision: 0.45 0.69  
recall: 0.10 0.94  
f1-score: 0.16 0.80  
support: 51 108  
QWK: 0.05

**llama3 - B2**



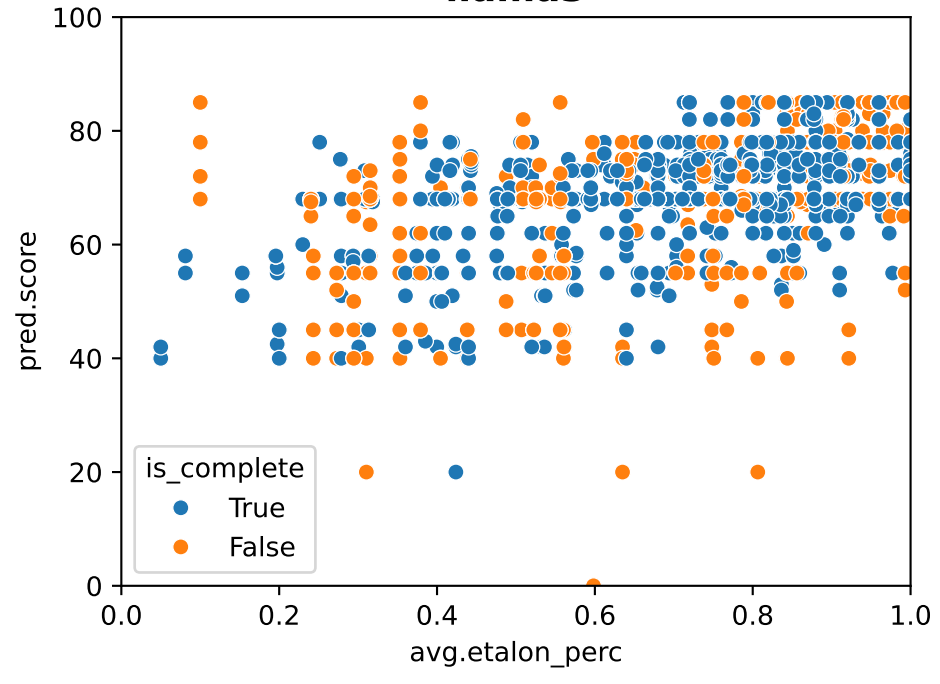
precision: 0.00 0.68  
recall: 0.00 0.99  
f1-score: 0.00 0.80  
support: 51 108  
QWK: -0.01

**deepseek-r1-70b**



precision: 0.32 0.82  
recall: 0.21 0.89  
f1-score: 0.25 0.86  
support: 234 969  
QWK: 0.12

**llama3**



precision: 0.50 0.83  
recall: 0.18 0.96  
f1-score: 0.27 0.89  
support: 234 969  
QWK: 0.18