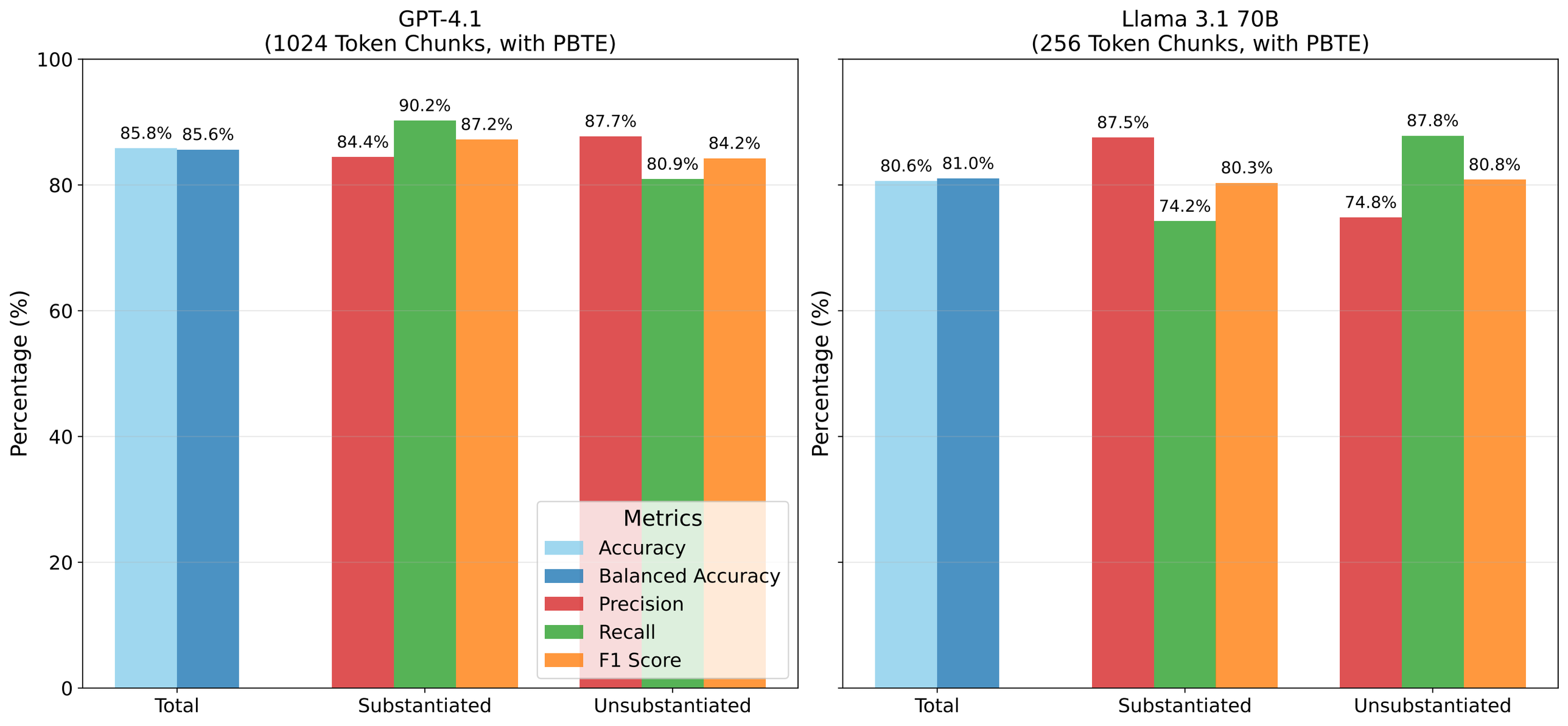
GPT-4.1
(1024 Token Chunks, with PBTE)
Llama 3.1 70B
(256 Token Chunks, with PBTE)
Percentage (%)
100
80
60
40
20
0
85.8%
85.6%
84.4%
90.2%
87.2%
87.7%
80.9%
84.2%
80.6%
81.0%
87.5%
74.2%
80.3%
74.8%
87.8%
80.8%
Total
Substantiated
Unsubstantiated
Metrics
Accuracy
Balanced Accuracy
Precision
Recall
F1 Score