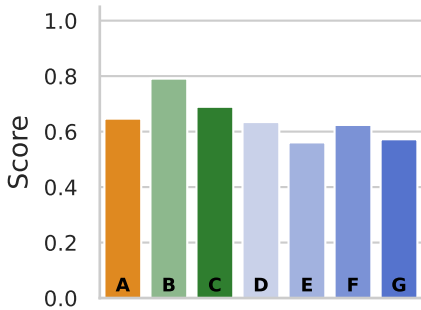
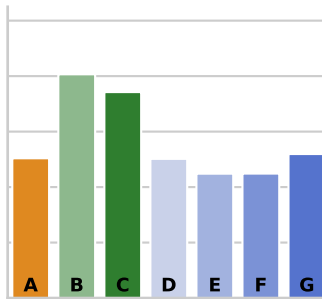


# Model Performance Comparison Across Metrics (Temp 0.0)

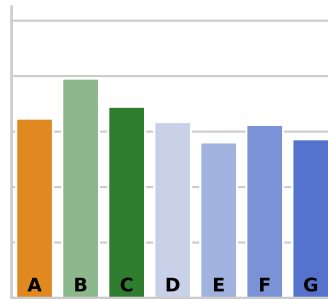
Metric: accuracy



Metric: precision



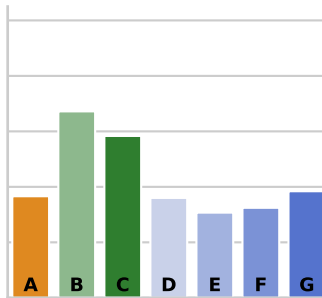
Metric: recall



Metric: f1\_score



Metric: balanced\_accuracy



## Model Key

A: Mistral-v0.3-7B  
B: gpt-4o  
C: gpt-4o-mini

D: llama-3.1-70B  
E: llama-3.1-8B

F: llama-3.2-3B  
G: llama-3.3-70B