F1 Score Comparison on both datasets 1.0 0 0 0 Score Model 0 string-matching 0.5 gpt-3.5-turbo 0 0 0 gpt-4o gpt-4.5-preview 0.0 1.0 $\stackrel{\circ}{\blacksquare}$ 0 Score 0.5 0 0 0 0 0 0 0 0.0 Feature Eng. Mep Geu.

Task Type