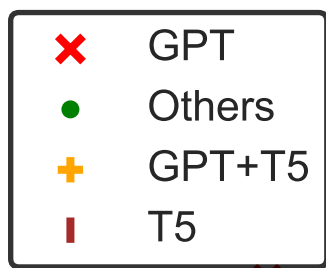


Real Queries, GPT-4 Labels



Kendall $\tau = 0.9151$

Real Queries, Human Labels