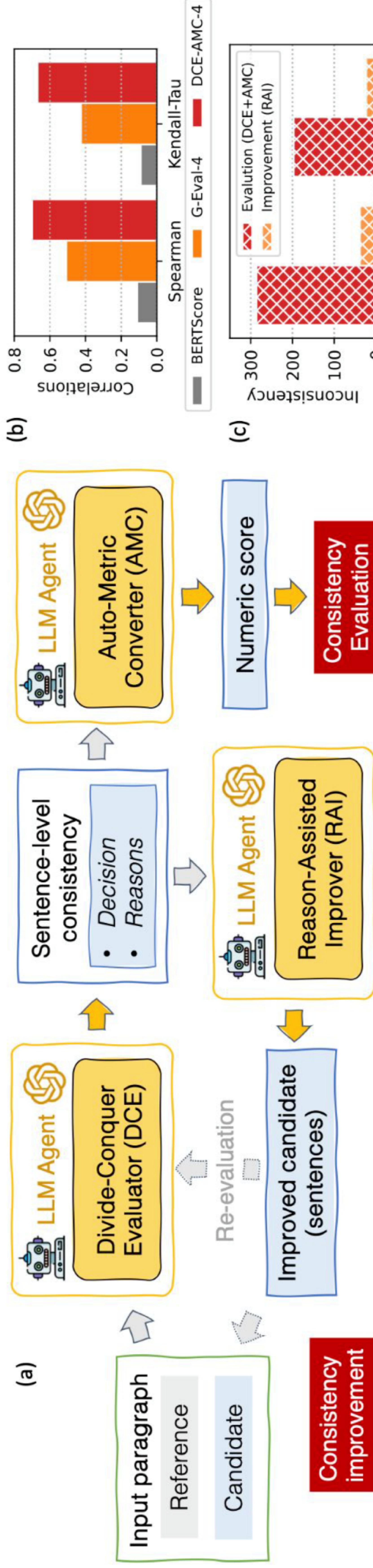


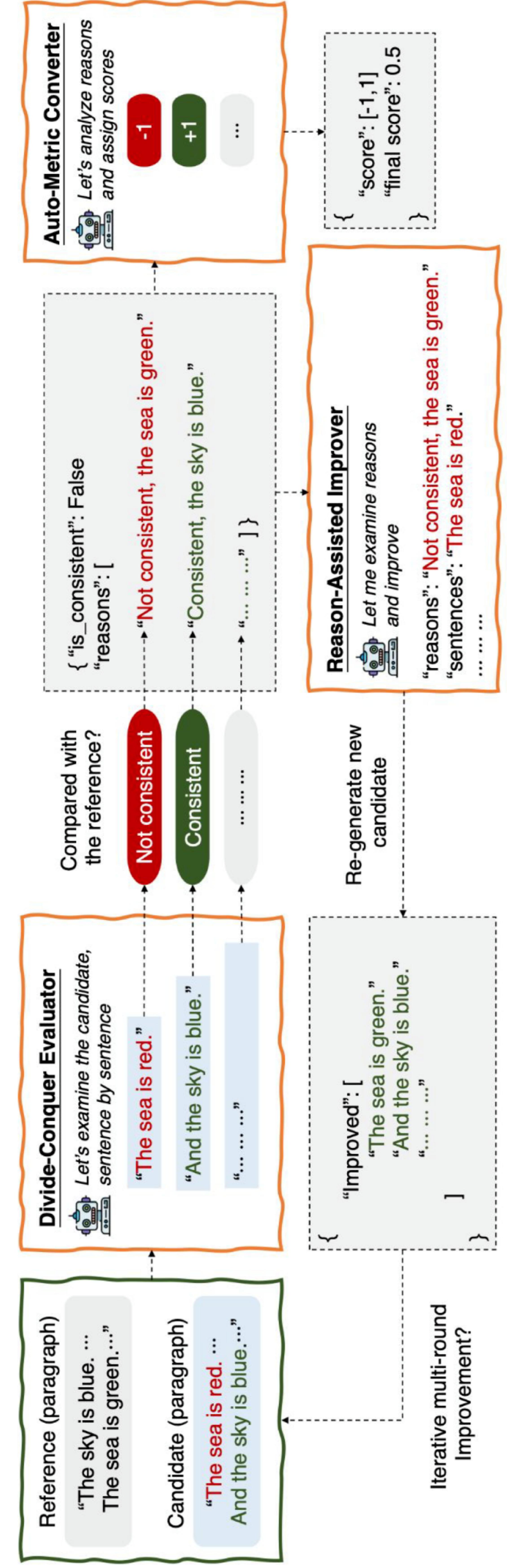
# Evaluate + Improve Generation Consistency via Divide-Conquer-Reasoning

Wendi Cui, Jiaxin Zhang, Zhuohang Li, Damien Lopez, Kamalika Das, Bradley Malin, Sricharan Kumar



(a) Overview of the proposed DCR framework. The first two components (DCE-AMC) aim at providing a better strategy for evaluating and quantifying semantic consistency to best match human judgments. Building on this, a third component RAI further utilizes analytical reasoning to iteratively improve the consistency of LLM-generated content w.r.t. the reference by minimizing hallucinations. (b) The combination of DCE and AMC (DCE-AMC-4) significantly outperforms the baseline methods in terms of correlations with human ratings. (c) RAI is able to mitigate inconsistencies in few rounds.

## example



## performance

