**GPT-2 Performance on Parasitic Gaps Across Stimulus Datasets** 100 **Evaluation Metric**  $\Delta$  filler > 0 Accuracy **DiD Accuracy** 80.0% 80 72.9% 68.8% 60.0% Accuracy (%) 60 Chance 40 20 7.0% 5.6% 0.0% 0 Original Lan et al. Filtered Lan et al. Refined Novel Data **Stimulus Dataset**