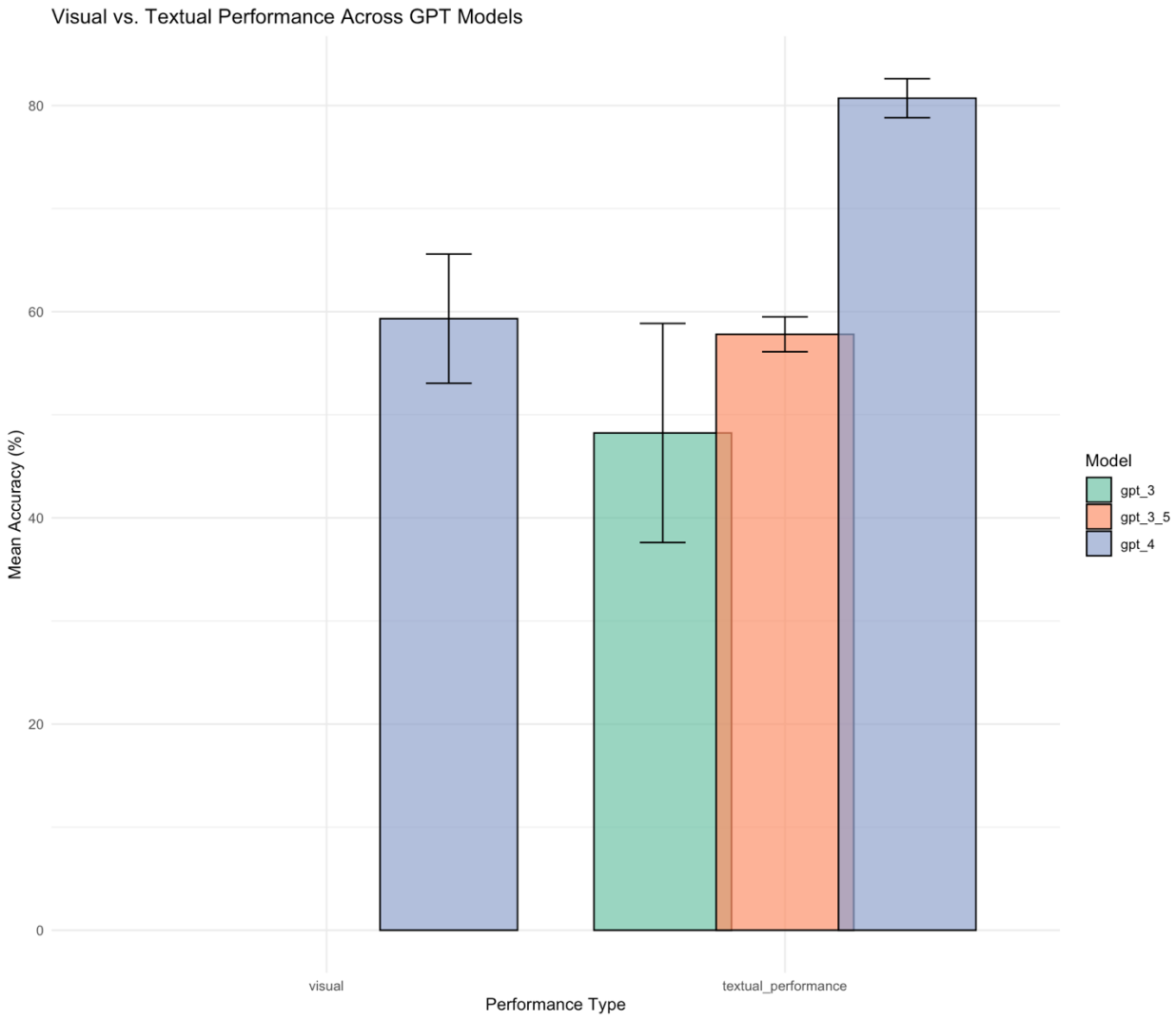

Supplemental Figure IV: Total Accuracy for Each GPT Model on Visual versus Text-based Questions

Total score is defined as the total number of correct questions divided by the total number of questions as reported by each study of interest. Visual performance was reported by n = 2 studies whereas textual performance was reported by n = 14 studies. ChatGPT-3 and ChatGPT-3.5 do not have visual recognition capabilities. The bars represent 95% confidence intervals.