Multi-Step Pipeline Tasks
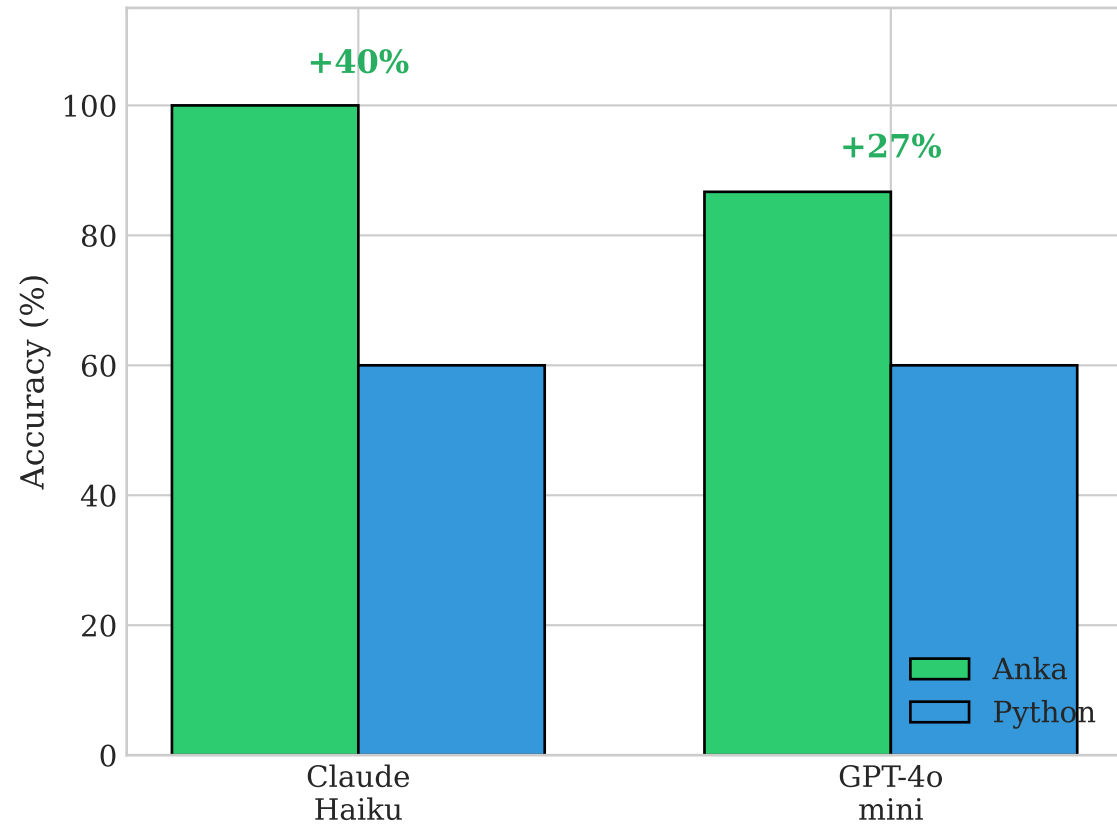
Overall Accuracy (Claude 3.5 Haiku)