## Swarm Plot for Task2 - Adjusted Scores by Rating Category 2 · Adjusted Score -2 -**-4** Rating Category

Model Name

cot-gpt-4-manual-cot cot-gpt-4-zero-shot-cot

babyagi-gpt-4

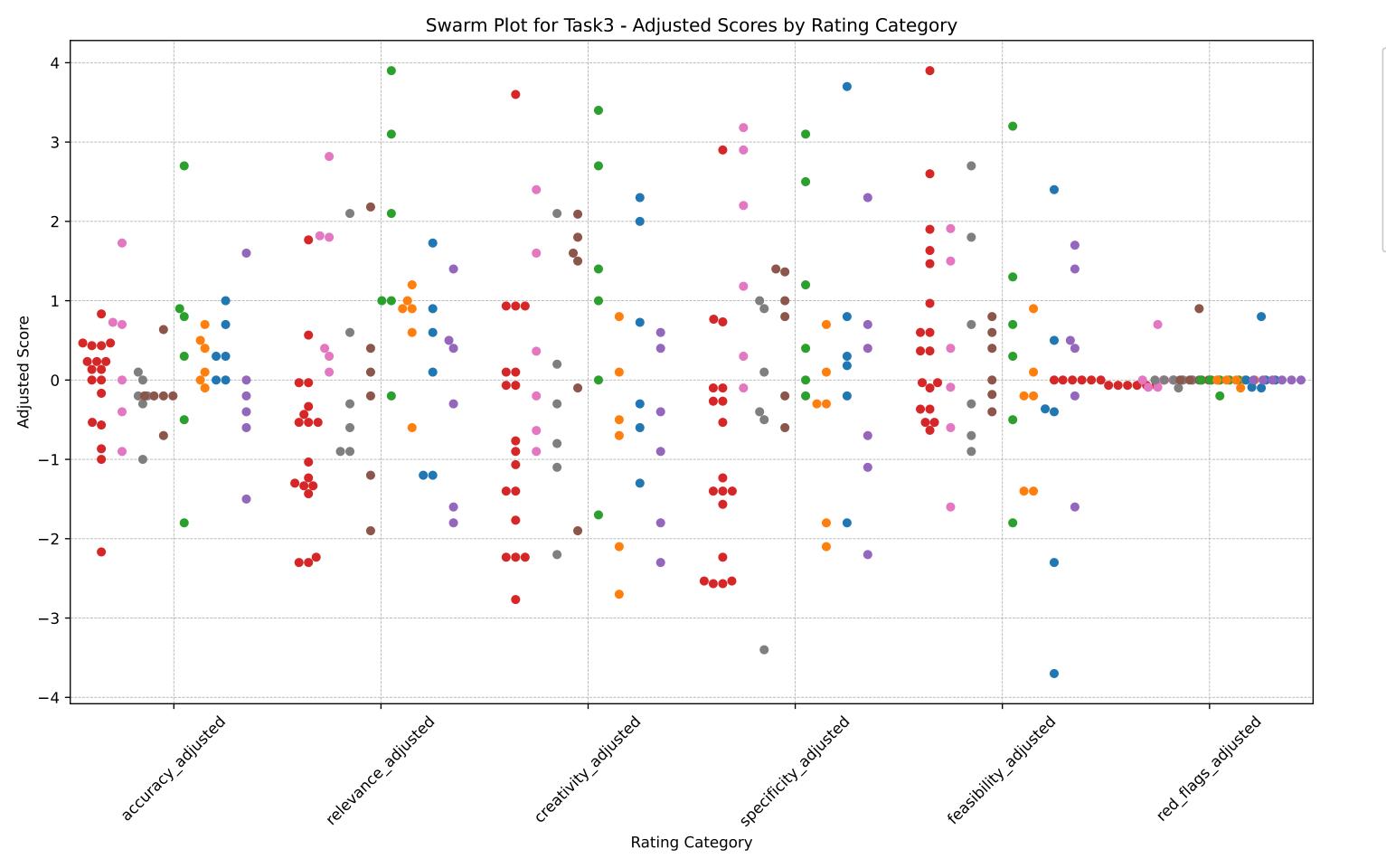
camel-gpt4

gpt-3.5-turbo

llama-2-7b-chat

bard

gpt-4



- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat

## Swarm Plot for Task1 - Adjusted Scores by Rating Category 6 Adjusted Score -2 · -4

Rating Category

Model Name

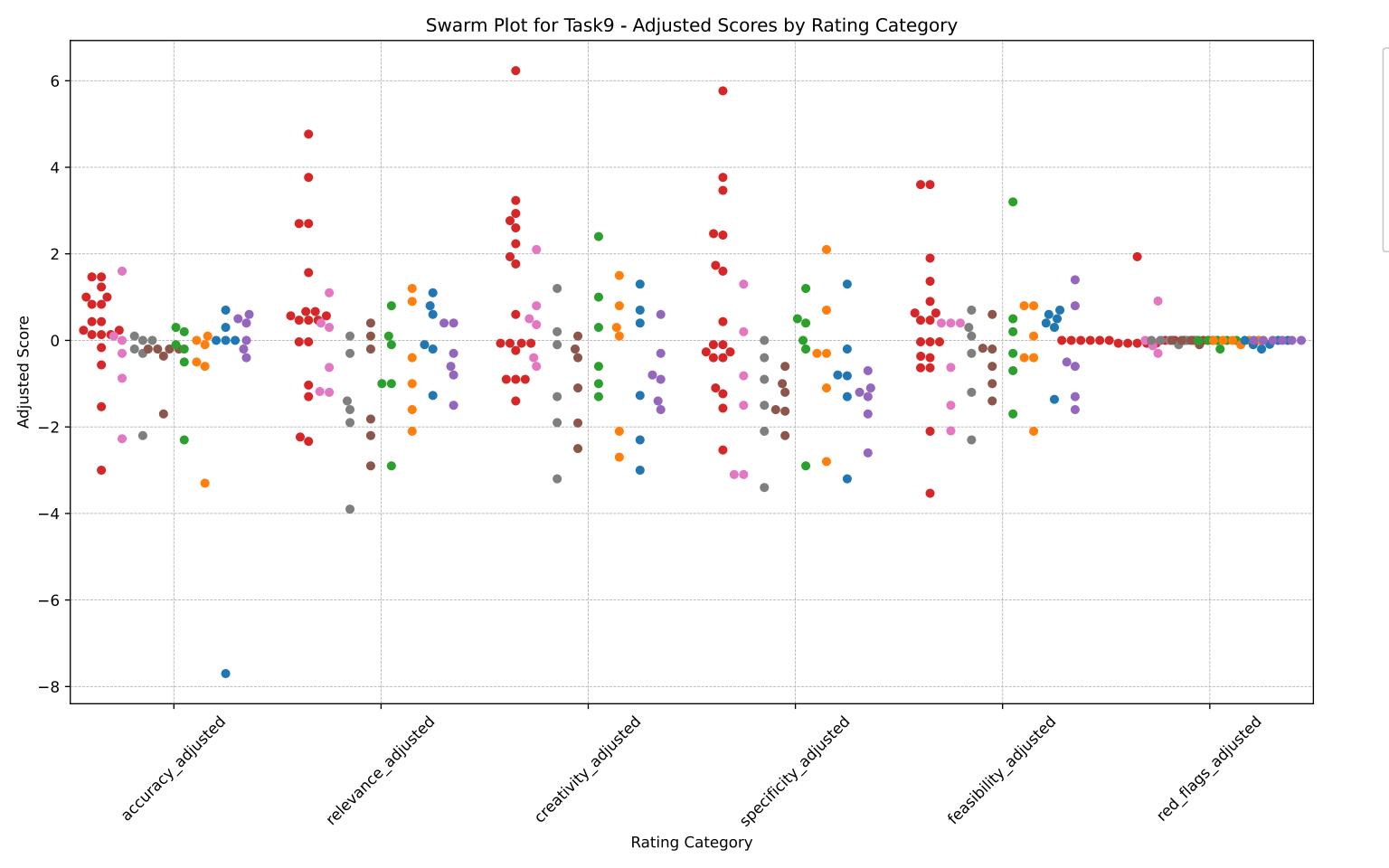
- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat

Swarm Plot for Task8 - Adjusted Scores by Rating Category 2 -Adjusted Score **-2** Rating Category

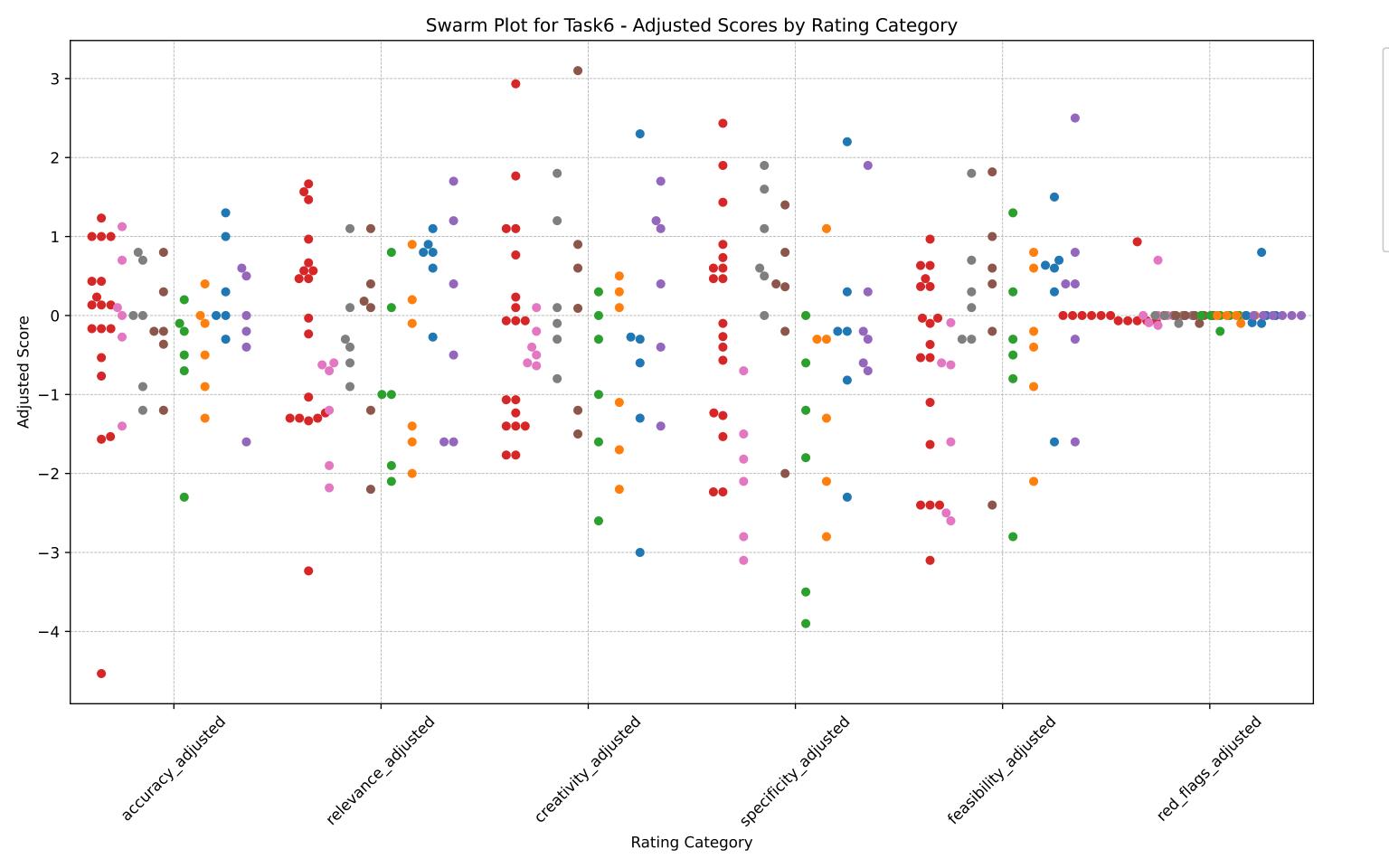
- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat

## Swarm Plot for Task10 - Adjusted Scores by Rating Category 3 2 -Adjusted Score **-2** --3 •• -4Rating Category

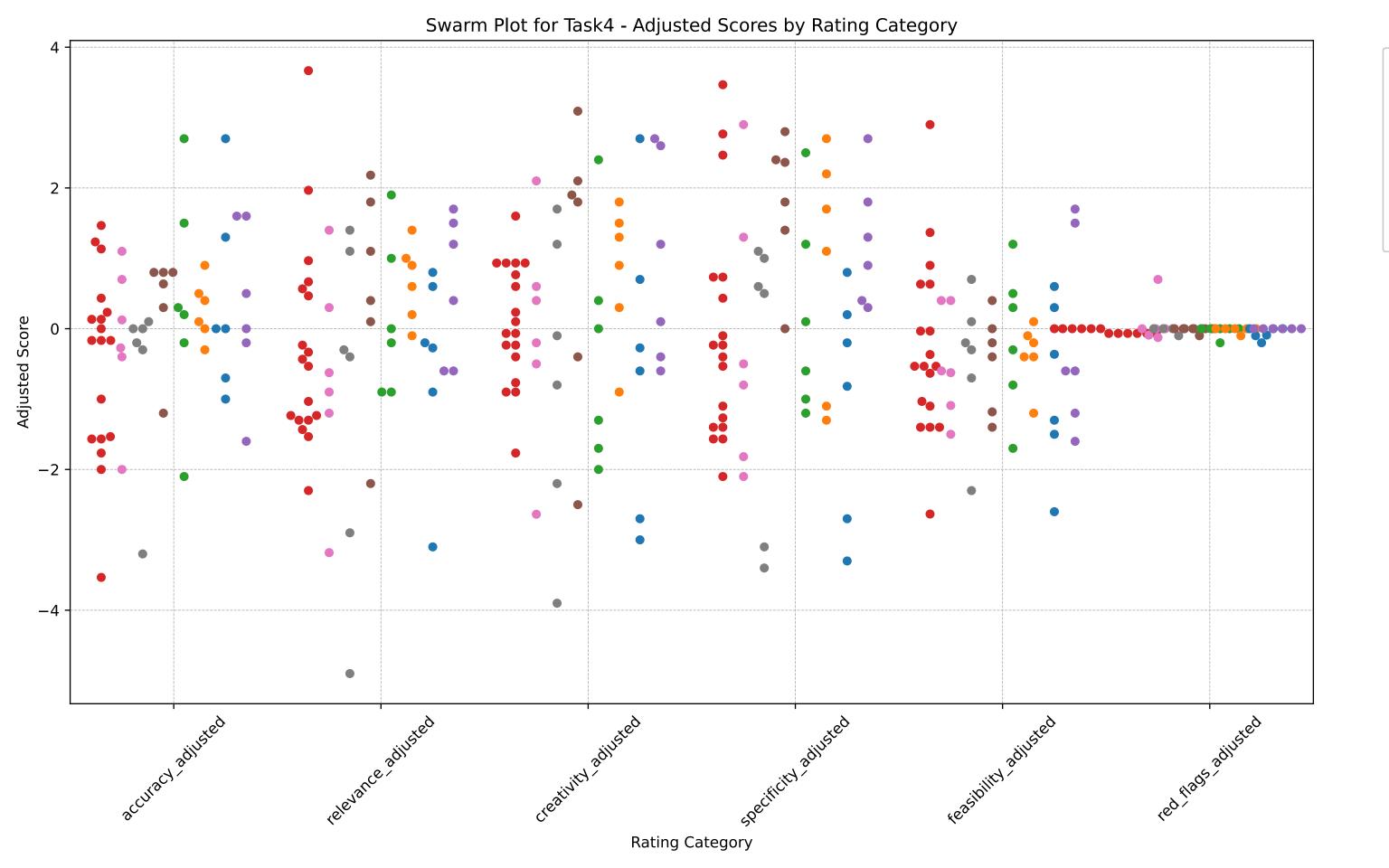
- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat



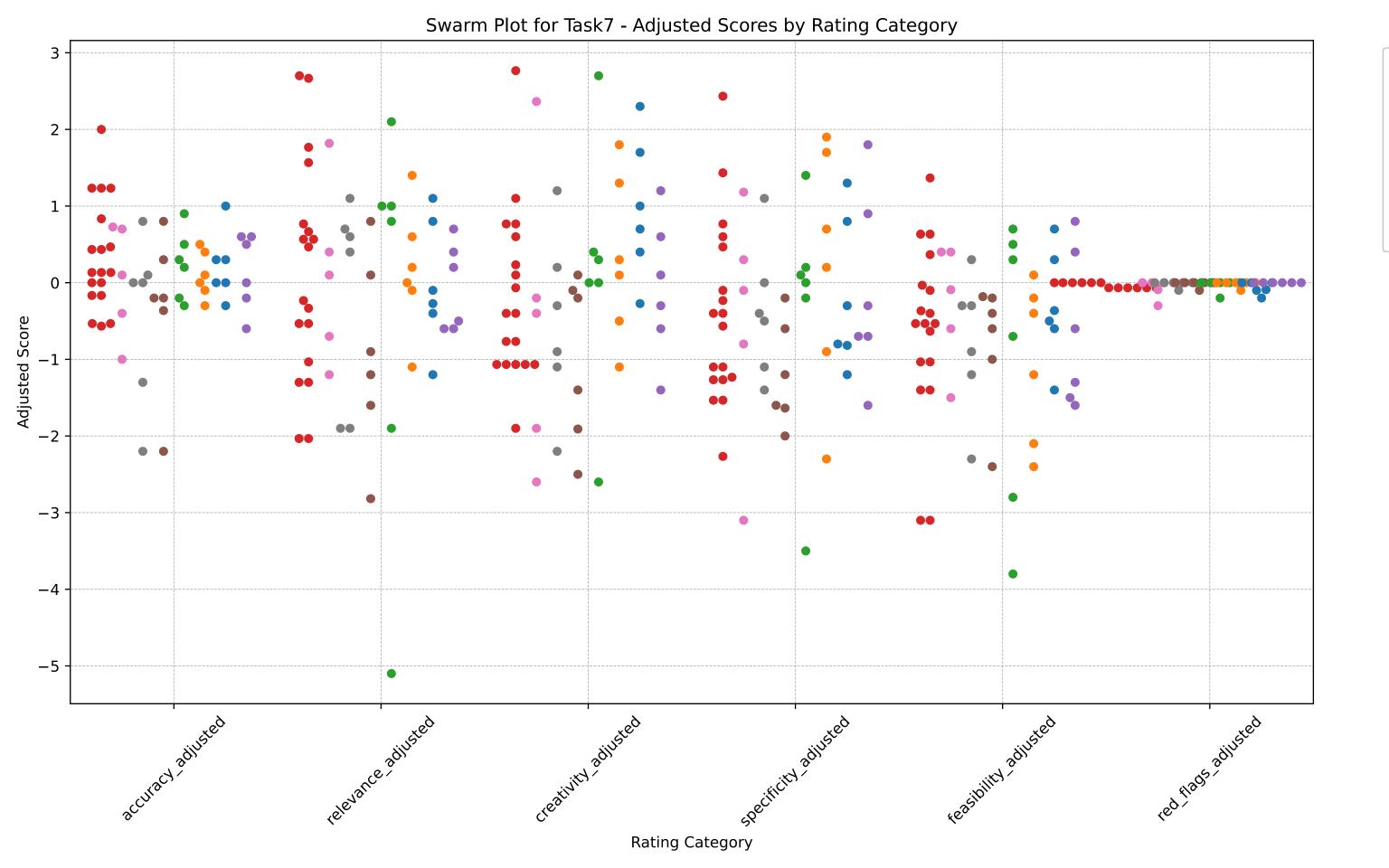
- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat



- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat



- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat



- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat

Swarm Plot for Task5 - Adjusted Scores by Rating Category 2 · Adjusted Score **-2** -**-4** Rating Category

- babyagi-gpt-4
- bard
- camel-gpt4
- cot-gpt-4-manual-cot
- cot-gpt-4-zero-shot-cot
- gpt-3.5-turbo
- gpt-4
- llama-2-7b-chat