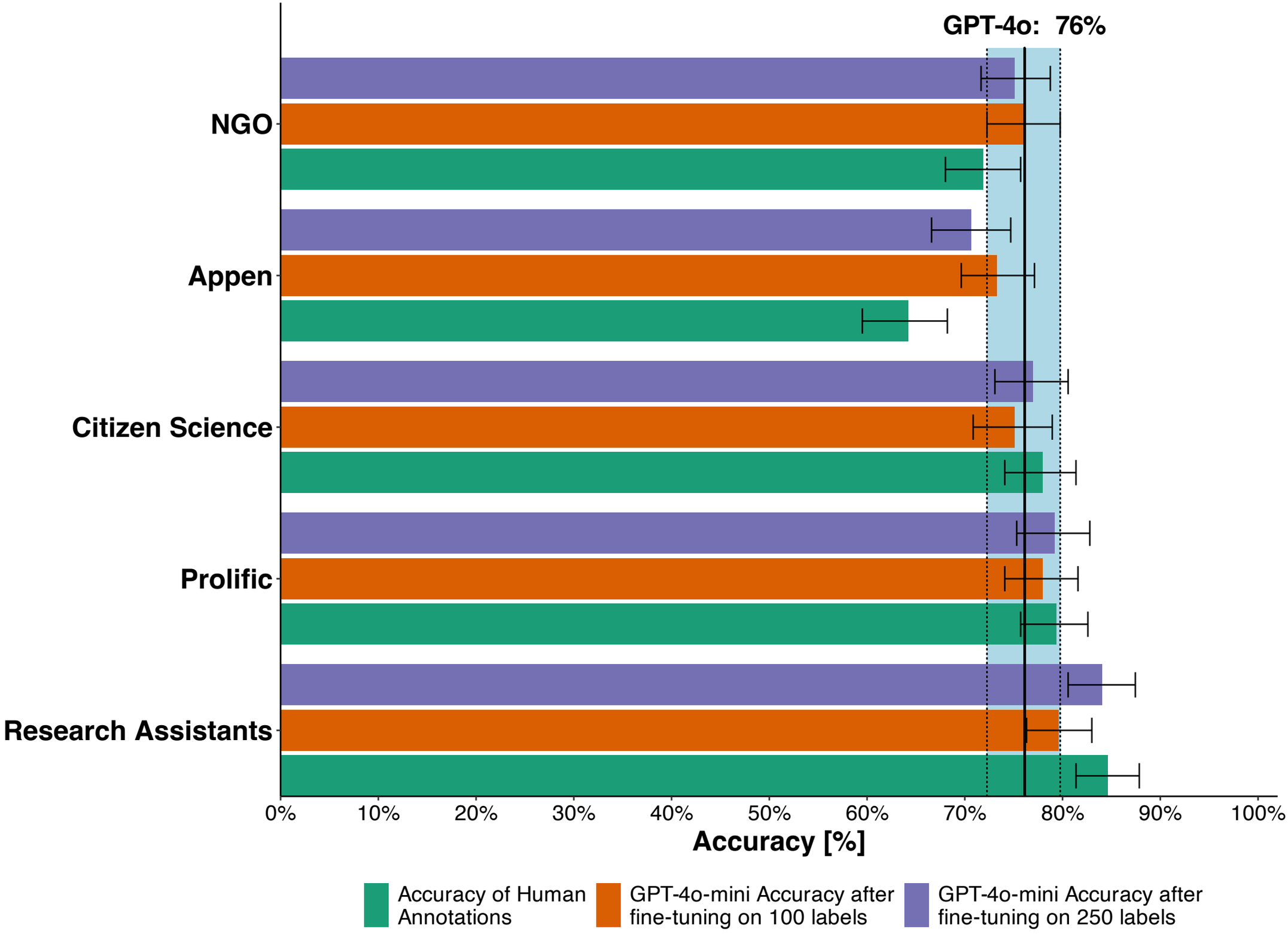
GPT-4o: 76%
NGO
Appen
Citizen Science
Prolific
Research Assistants
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Accuracy [%]
Accuracy of Human Annotations
GPT-4o-mini Accuracy after fine-tuning on 100 labels
GPT-4o-mini Accuracy after fine-tuning on 250 labels