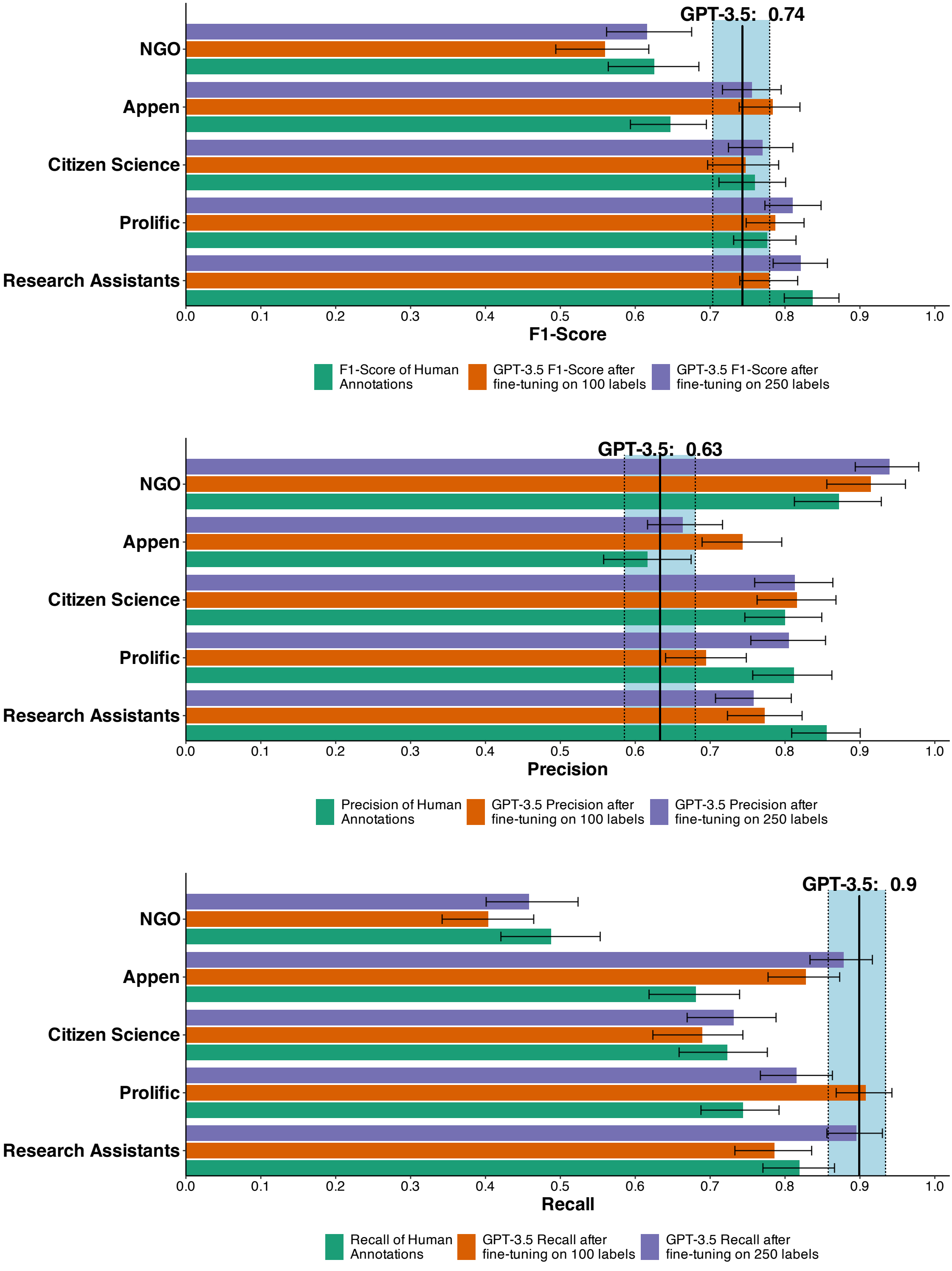
GPT-3.5: 0.74
NGO
Appen
Citizen Science
Prolific
Research Assistants
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
F1-Score
F1-Score of Human Annotations
GPT-3.5 F1-Score after fine-tuning on 100 labels
GPT-3.5 F1-Score after fine-tuning on 250 labels
GPT-3.5: 0.63
Precision
Precision of Human Annotations
GPT-3.5 Precision after fine-tuning on 100 labels
GPT-3.5 Precision after fine-tuning on 250 labels
GPT-3.5: 0.9
Recall
Recall of Human Annotations
GPT-3.5 Recall after fine-tuning on 100 labels
GPT-3.5 Recall after fine-tuning on 250 labels