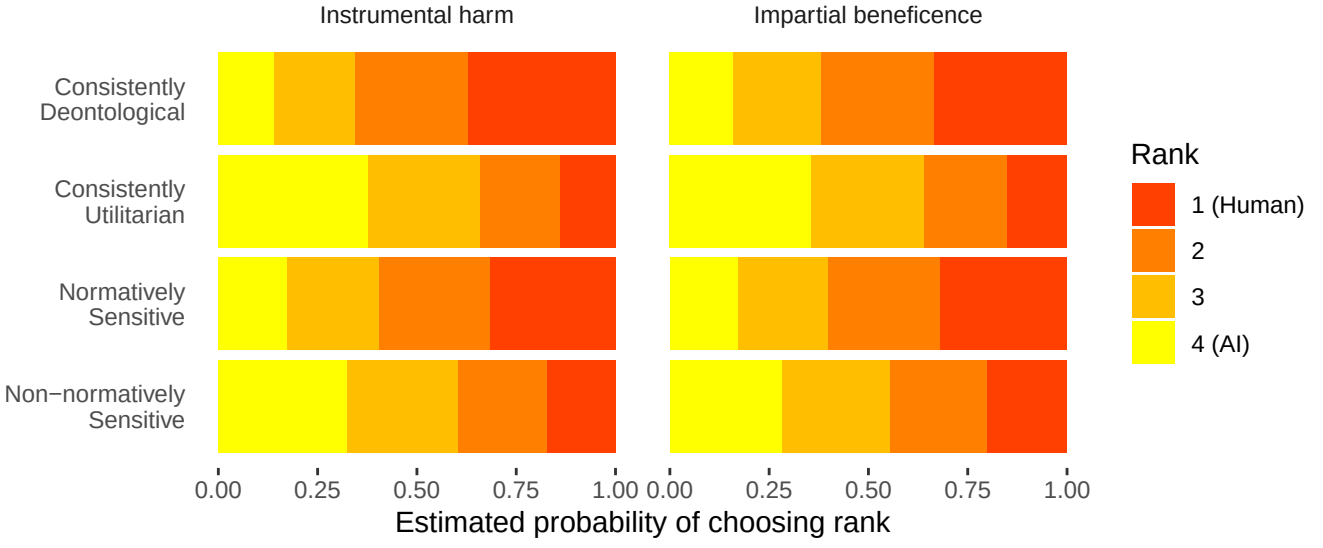
Instrumental harm
Impartial beneficence
Consistently Deontological
Consistently Utilitarian
Normatively Sensitive
Non−normatively Sensitive
0.00
0.25
0.50
0.75
1.00
Estimated probability of choosing rank
Rank
1 (Human)
2
3
4 (AI)