

Advisor

Human

AI

Advice



Deontological



Utilitarian

-0.2

-0.1

0.0

0.1

0.2

Pre-post shift in judgement after seeing advice (0-1 scale)

Deontological \leftrightarrow Utilitarian

