

В.Э. Карпов

О моральных аспектах адаптивного поведения искусственных агентов

В докладе обсуждается перечень свойств, которыми должен обладать моральный агент и показывается, что искусственный агент не только может обладать такими свойствами, но эти свойства могут быть обусловлены с технической точки зрения, как проявления адаптационных механизмов.

В частности показано, что такие механизмы, как наличие компоненты “Я” в знаковой картине мира агента, наличие эмоционально-потребностной архитектуры, а также механизм сопоставления наблюдаемого контрагента с “Я” позволяют реализовать феномены социального обучения и такого свойства, как эмпатия. Приводится описание экспериментов, основным варьируемым параметром которых была склонность искусственных агентов к симпатии.

Задачей экспериментов было определение зависимости показателей "благополучия" группы агентов от степени их моральности – готовности к альтруизму.