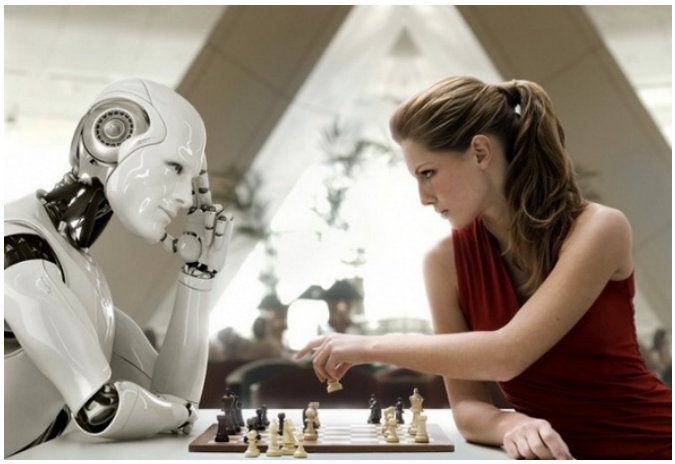
Автор: Administrator
17.08.2012 13:56 -



Глобальная катастрофа определена Бостромом (Бостром – ученый-трансгуманист) как такая, что полностью истребит разумную жизнь на земле или необратимо повредит ее потенциал. Юдковски предлагает разделить потенциальные ошибки в попытках создать дружественный интеллект на две категории: ошибку техническую и ошибку философскую. Техническая ошибка заключается в том, что созданный искусственный интеллект не будет работать так, как должен. А философская – не получится облагодетельсьтвовать все человечество, даже создав искусственный интеллект.

Примером философской ошибки может служить стремление порядочных и образованных людей 19 века к воплощению идеи коммунизма. Эти люди были идеалистами. Они не могли предсказать, что произойдет, когда в России победила революция. Идеалисты предполагали, что люди не будут вынуждены работать много часов, получая за это гроши. Они ошиблись. Сделать всех людей счастливыми не удалось. А если вообразить, что кто-то запрограммирует дружественный искусственный интеллект на внедрение его любимой политической системы. Ему это будет казаться предвестником всеобщего счастья. То есть, программист доверяет себе настолько, что не верит в вероятность собственной ошибки. Это программистская ошибка на моральном уровне. То есть, выбор в пользу коммунизма, например, происходит из-за эмпирической веры плюс ценностное суждение. Эмпирическая вера говорит, что коммунизм сделает всех счастливее (меньше работать и быть богаче), ценностное суждение заключается в том, что такой результат – это то, что нужно сейчас большинству людей.

Возможно, оправданны ожидания, что искусственный интеллект сможет менять свои эмпирические верования. Если бы искусственный интеллект появился до открытий Коперника и изобретения телескопа, то он верил бы, что Солнце вращается вокруг Земли. А позднее ему пришлось бы поменять свое мировоззрение.