Автор: Administrator
08.08.2012 00:00 -

Создав сильный искусственный интеллект, мы можем создать собственную надежду на выживание – искусственный интеллект может глобально анализировать будущие риски для человечества. Первыми установками при создании искусственного интеллекта должны стать сильные установки для дружественности. Они не должны оказаться недоступными, когда у людей появится возможность создавать мощные искусственные интеллекты.

Очевидно, что задача, которую человек может решить, пропорциональна силе его интеллекта. Значит, сильный искусственный интеллект сможет решать сверхзадачи. Вопрос в том, захочет ли он это делать. Говоря по-другому, будет ли у него для этого мотив.

Распространенные стереотипы, в связи с искусственным интеллектом:

- мы должны отказаться от идеи создания искусственного интеллекта потому, что он уничтожит человечество; (и он захочет это сделать)

- нам нужен искусственный интеллект, потому что он способен изобрести лекарства от многих болезней и спасти человечество; (и он захочет это сделать)

- способности искусственного интеллекта будут гораздо выше человеческих во всех значимых областях, он будет выполнять всю работу, а человечество будет бездельничать (и он захочет работать).

Эти заблуждения рождены антропоморфизмом (видением искусственного интеллекта через призму собственных способностей и мотивов). Мы должны научиться управлять будущим, направляя его в область, полезную для человечества.

Вывод: человечеству нужен хороший оптимизационный процесс для прогнозирования результата создания искусственного интеллекта. То есть, человечеству нужен искусственный интеллект с заданной специфической мотивацией. Многие скептически заявляют, что сильный искусственный интеллект сможет разорвать всякие наложенные на него обязательства – каждый искусственный интеллект сможет изменить свой исходный код. Но нормальный, обладающий интеллектом человек не станет принимать пилюлю, заставляющую его наслаждаться убийством, например, мы не хотим, чтобы люди умирали.