



ИСКУССТВЕННЫЕ ЭГО-МАШИНЫ. ПРОБЛЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В НЕЙРОФИЛОСОФИИ.

Доклад магистранта СПбГУ
Снеткова Ивана
Геннадьевича

Контактные данные:
i.snetkov@yandex.ru



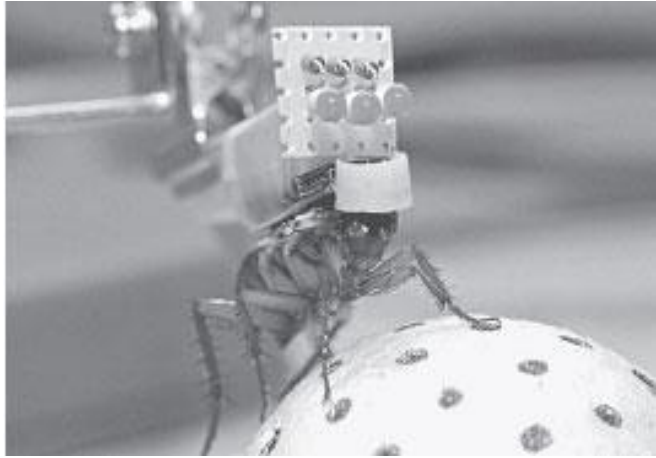
Понятие Эго-машины.

Эго-машина – система, способная производить осознаваемое Я.

Эго-машина необязательно должна быть живым существом: она может быть чем угодно, лишь бы обладала осознаваемой я-моделью.

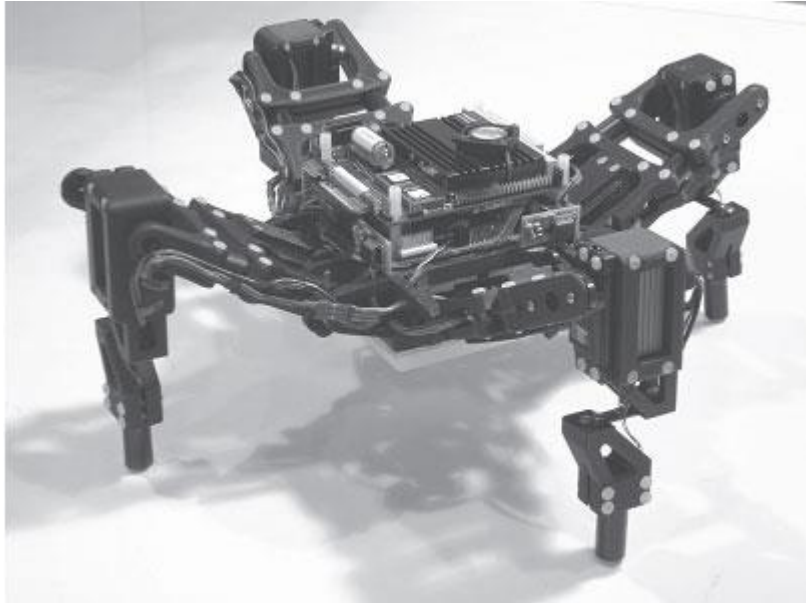
Концептуальное различие между искусственными и естественными системами не исчерпывающее и не эксклюзивное: то есть возможны разумные и/или сознающие системы, не принадлежащие ни к одной из этих категорий.

Системы, использующие биологическое «железо», управляемое искусственной, (то есть созданной человеком) программой.



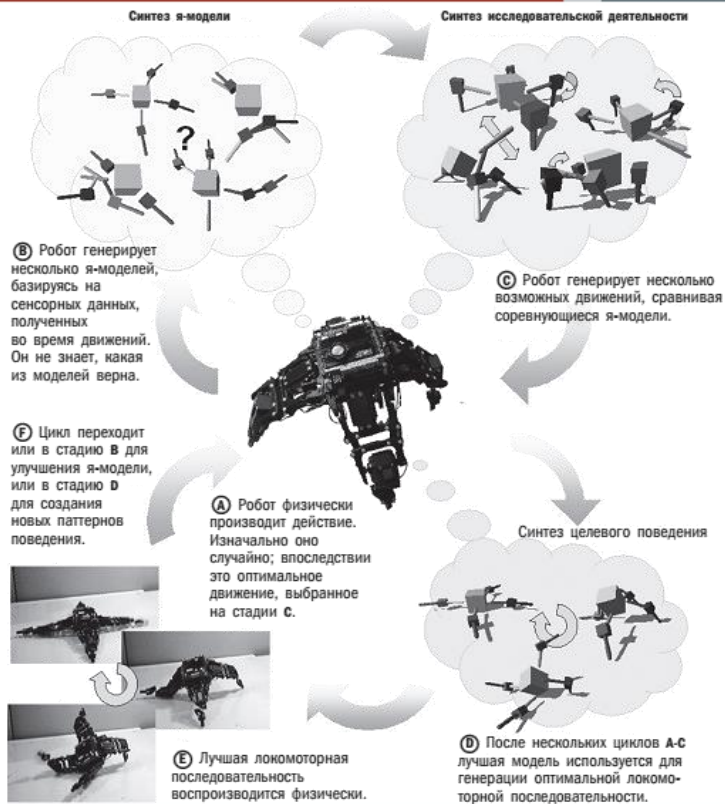
RoboRoach, Роботаракан. Управление движением таракана через вживленный ранец-микроробот. Тараканий «ранец» содержит приемник, преобразующий сигнал с удаленного пульта в электрический стимул, подаваемый на антенны таракана. Это позволяет оператору приказать таракану остановиться, двигаться вперед, пятиться и поворачивать вправо или влево. На сегодняшний день возможно успешно управлять также крысами или летающими насекомыми, посылая сигнал напрямую в нервную систему. Фото: Associated Press.

Искусственное «железо», управляемое естественно развившимися программами.



Морская звезда – четырехногий робот, способный ходить, используя внутреннюю модель себя, которую он создал и непрерывно совершенствует. Утратив ногу, он способен внести поправку во внутреннюю я-модель.
Фото: Джош Бонгард

Циклы действий робота во время движения.



А и В: Синтез я-модели. Робот физически производит действие.

А: Сначала действие производится случайным образом, затем выбирается лучшее из действий, найденное в стадии С.

В: Затем робот генерирует несколько я-моделей, сравнивая сенсорные данные, собранные при выполнении предшествующего движения. Он знает, которая из моделей верна.

С: Синтез экспериментальных данных. Робот генерирует несколько возможных движений, сравнивая конкурирующие я-модели.

Д: Синтез целевого поведения. После нескольких циклов от А до С лучшая из наличных моделей используется для генерации оптимальной локомоторной последовательности.

Е: Лучшая из локомоторных последовательностей выполняется физически. Фото: Джош Бонгард

Я-модель эго-машины.



Я-модели представляют собой изобретение естественной эволюции.

Я-модель может быть неосознанной, она может развиваться, и она может быть создана в машине, подражающей процессу биологической эволюции.

Суммируя: у нас уже есть машины не исключительно естественные и не исключительно искусственные.

Вероятно, что первая не полностью биологическая форма сознательного чувства Я впервые будет реализована в таких эго-машинах.

Как построить искусственного сознающего субъекта, и почему этого не следует делать.



Для этого необходимо ответить на следующие вопросы:

1. При каких обстоятельствах можно обоснованно предположить, что искусственный интеллект имеет осознаваемое «Я» и действительно осознанно переживаемым взглядом от первого лица?
2. Что превращает искусственную систему обработки информации в переживающий субъект?
3. Какие этические проблемы возникают, при появлении подобного искусственного переживающего субъекта?
4. Можно ли считать сознательный опыт как таковой благом?

Как построить искусственного сознающего субъекта, и почему этого не следует делать.



Ответы, предлагаемые нейрофилософом Т. Метцингером:

1. Сознательная система – это система, оперирующая глобально доступной информацией посредством единой внутренней модели реальности.
2. Внутренняя суть сознания *в присутствии мира*. Эго-машина, чтобы мир ей явился, должна обладать еще двумя функциональными характеристиками:
 - 1) переживаемое Сейчас, 2) транспарентность внутреннего образа себя.
3. Эго-машина может страдать, поскольку интегрирует болевые сигналы, состояния эмоционального расстройства и негативные мысли в свою транспарентную Я-модель, а значит эго-машина становится полноценным субъектом этики.
4. Появление эго-машины, означает появление не просто сознающего, но страдающего субъекта, а значит и увеличение количества страдания в мире, а наша задача предотвращать и уменьшать страдания всюду, где это возможно. У эго-машин будут собственные интересы, и этот факт они будут осознанно переживать, не имея никаких правовых институтов в защиту.

Как построить искусственного сознающего субъекта, и почему этого не следует делать.

Пока такие вопросы не нашли ответов, конвенционально признанных в мировом научном сообществе, нам следует воздержаться от всяких попыток создания искусственных сознающих субъектов.

Необходимо помнить о необратимости некоторых ступеней развития. Любой искусственный интеллект, который будет близок к обретению характеристик сознающего субъекта, обладающим феноменальным чувством «Я», – станет одновременно самостоятельным действующим лицом с собственными предпочтениями, а значит и волевым, этическим, правовым, рациональным агентом и т. п.

Вместе с тем, необходимо продолжать исследования на предмет того, как возможно существование искусственного интеллекта, как сознающего субъекта, прежде всего для того, чтобы не допустить случайного создания подобного субъекта.



1. Анохин К. Выступление на OpenTalks.AI. Открытая конференция по искусственному интеллекту, 2018 [Электронный ресурс] URL: <https://www.youtube.com/watch?v=YMg5r95MS7M> (дата обращения: 18.02.2020)
2. Метцингер Т. Наука о мозге и миф о своем Я. Тоннель эго/Т. Метцингер. — Москва: Издательство АСТ, 2017
3. Нейрофилософия: вопросы, ожидания и перспективы // Философские науки. — 2015 - № 11
4. Bongard J. et al. Resilient machines through continuous self-modeling // Science 314:1118–21 (2006)
5. Metzinger, Thomas. Being No One: The Self-Model Theory of Subjectivity. Cambridge, MA: MIT Press, 2003
6. Metzinger T. Two principles for robot ethics /: E. Hilgendorf & J.-P. Gxntner (eds.), Robotik und Gesetzgebun. Baden-Baden, 2013 Стр. 247–286
7. Metzinger T. The Myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy // Front. Psychol. 4:619 (2013)