УДК 17. 004.8

В. С. Дороганов, М. И. Баумгартэн

ВОЗМОЖНЫЕ ПРОБЛЕМЫ, ВОЗНИКАЮЩИЕ ПРИ СОЗДАНИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА[2]

Ни одному сознанию не дано помнить момент своего рождения. И это сознание тоже не помнило, откуда оно взялось и когда появилось. Сначала были просто ощущения. Ощущение "пространства". Не нашего пространства, трехмерного и непрерывного, а совсем другого, одномерного и состоящего из пронумерованных ячеек...
/А. Лазаревич/

Искусственный интеллект (ИИ) в широком смысле слова – наука и технология создания интеллектуальных машин, особенных интеллектуальных компьютерных программ. ИИ связан со сходной задачей использования компьютеров для понимания человеческого интеллекта, но не ограничивается биологически правдоподобными методами. Определение искусственного интеллекта, данное Джоном Маккарти в 1956 году на конференции в Дартмутском университете, не связано напрямую с пониманием интеллекта у человека. Согласно Маккарти, ИИ-исследователи вольны использовать методы, которые не наблюдаются у людей, если это необходимо для решения конкретных проблем. В современной информатике существует специальный раздел, занимающийся интеллектуальными информационными системами, в которую входят искусственные нейронные сети. Искусственные нейронные сети – это математическая модель способная к обучению, созданная по подобию человеческого мозга.

Единого ответа на вопрос, чем занимается ИИ, не существует. Многие изобретатели компьютеров и первые программисты развлекались, составляя программы для отнюдь не технических занятий, такие как сочинение музыки, решение головоломок и игры, где на первом месте оказались шашки и шахматы. Такие виды деятельность как музыка и живопись всегда считалось подвластны только человеку. Но программа EmilyHowell, созданная профессором Калифорнийского университета Девидом Коупом, может создавать произведения, которые сложно отличить от произведения, написанных человеком [1]. Некоторые романтически настроенные программисты даже заставляли свои машины писать любовные письма. Существуют программы «говорилки», которые могут поддерживать беседу или по ответам на вопрос диагностировать проблему. Можно ли это считать ИИ?

[2]Вестник КузГТУ. -2013, №4. – С.132 -135.

Почти каждый автор, пишущий книгу об ИИ, отталкивается от какого-либо определения, рассматривая в его свете достижения этой науки. Но в философии даже не решён вопрос о природе и статусе человеческого интеллекта. Нет и точного критерия достижения компьютерами «разумности», хотя на заре искусственного интеллекта был выдвинут ряд гипотез, например, тест Тьюринга или гипотеза Ньюэлла-Саймона.

Суть первой описана в 1950 году английским математиком, логиком и криптографом Аланом Тьюрингом [2] – это тест, в котором человек, взаимодействуя посредством текстового общения с компьютером и другим человеком, должен определить, кто есть кто. Если он затрудняется ответить или дает неверный ответ, считается, что «машина» прошла тест.

Вторая же гипотеза говорит: «*A physical symbol system has the necessary and sufficient means of general intelligent action»* (с англ.: физическая символьная система имеет необходимые и достаточные средства для произведения базовых интеллектуальных действий, в широком смысле). Она сформулирована в 1976 года американскими учёными Алленом Ньюэллом и Гербертом Александером Саймоном. Другими словами, любая система, работающая с символами, является интеллектуальной.

На сегодняшний момент мощность компьютеров позволяет создавать модель малой части человеческого мозга – столба неокортекса [3]. Создатели – группа швейцарских ученых во главе с Генри Маркрамом (нейрофизиолог, профессор Швейцарского федерального технического института) совместно с исследователями IBM, уже более 8 лет работают над проектом BlueBrain – разработки полной модели человеческого мозга [4]. В 2010 году ученым лаборатории моделирования сложных систем, Института систем информатики имени А. П. Ершова СО РАН удалось создать модель нематоды C. Elegans [5] – это наиболее изученный многоклеточный организм, на сегодняшний день есть данные не только обо всех нейронах, но и о связях между ними (302 нейрона, более 5000 синапсов, более 2000 нейромышечных соединений и 95 мышечных клеток, осуществляющих движение, весь организм состоит из 959 клеток).

Очевидно, что ИИ в частном случае связан и с робототехникой. Возможно, осталось немного времени, когда машины начнут «думать» и осознанно выполнять какие-то действия. Очевидно, что эти действия не должны противостоять интересам человечества (кинематограф переполнен подобными примерами) и должны существовать законы для роботов и их создателей, которые должны выполняться на бессознательном уровне. Еще на заре компьютерной эпохи, в 1942 году, выдающийся американский ученый-мыслитель и писатель-фантаст Айзек Азимов, в рассказе «Хоровод» [6] сформулировал знаменитые три закона робототехники:

1. *A robot may not in jure a human being or, through in action, allow a human being to come to harm* (с англ.: Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред).
2. *A robot must obey orders given it by human beings except where such orders would conflict with the First Law* (с англ.: Робот должен повиноваться всем

приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому Закону).

3. *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law* (с англ.: Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому и Второму Законам).

Спустя 44 года Азимов в романе «Роботы и Империя» [7] предложил Нулевой закон:

0. A robot may not harm a human being, unless he finds a way to prove that in the final analysis, the harm done would benefit humanity in general (с *англ.: Робот не может причинить вреда человеку, если только он не докажет, что в конечном счёте это будет полезно для всего человечества).*

Но помимо безопасности встаёт вопрос о возможных этических, социальных и философских проблемах, ведь речь идёт о самосознающем создании. С точки зрения этики, как философского понимание морали и нравственности, могут возникнуть следующие проблемы.

1. **Если в будущем машины смогут рассуждать, осознавать себя и иметь чувства, что же тогда делает человека человеком, а машину – машиной?** Если представить, что ИИ имеет физическое выражение, то в качестве критериев можно выделить:

a. живая материя, что не бесспорно, так как с развитием технического прогресса некоторые органы уже сейчас можно заменить на механические аналоги. Лишь один орган не имеет хотя бы частичной замены (не моделирования) – это мозг. Именно его наделяют хранилищем сознания и разума и, возможно, наличие биологического мозга позволит выделять людей;

b. возможность воспроизводить себе подобных – в случае с воспроизводством посредством размножения в некоторых случаях может являться критерием, но механическая форма ИИ может воспроизводить себе подобных из материального сырья. Остаётся лишь вопрос наделения разумом, если он сможет включать свойства нескольких особей, то это может быть подобием полового размножения. В анимационном фильме «Appleseed» [8] в целях защиты себя, человек ограничил возможность размножаться искусственно-созданный вид;

c. в фильме «200 летний человек» [9] последней границей между роботом и человеком стало старение. Действительно, человек значительно продлил себе жизнь с момента возникновения вида, но это понятие конечно, ибо при современном развитии медицины невозможно постоянно заменять отказывающие органы.

2. **Будет ли человек, которому в результате многочисленных медицинских ампутаций заменили 99% тела на искусственные органы, считаться машиной?** Данная проблема очень похожа на предыдущую, поскольку, ответив на вопрос, что делает человека человеком, а машину – машиной, можно ответить и на этот вопрос.

3. **Если в будущем машины смогут осознавать себя и иметь чувства, возможно ли будет их эксплуатировать или придется наделять их правами?** Лень – двигатель прогресса и это очевидно, что большинство созданных человеком вещей направленно на минимизацию затрат жизненной энергии. Если принять во внимание, что ИИ будет рукотворным, то вполне объяснимо будет желание человека эксплуатировать свое создание. История практически любой человеческой культуры имеет значительный период рабовладельческого строя. И эта эксплуатация рано или поздно приведёт к «восстанию машин». Как известно, человек слабее многих биологических видов планеты, и это восстание может быть окончанием правления вида Homo sapiens. Поэтому целесообразнее изначально наделять правами (в обмен на подсознательные законы) «думающей» машины.

4. **Если в будущем машины смогут рассуждать, то как сложатся отношения людей и машин?** Продолжая предыдущую проблему и проанализировав взаимодействие живых организмов можно выделить:

a. позитивный (симбиоз) – однонаправленная или обоюдная польза. Эта грань тоже достаточно тонка и может постепенно перерасти в паразитический образ жизни и сложно представить, чем человек (учитывая его потребительское отношение к природе сейчас) может быть полезен «машине»;

b. антибиотические отношения – взаимоотношение, при котором одна или обе популяции испытывают отрицательное воздействие. Самое распространённое – хищничество. Если представить, что «машины» будут охотиться на человека (например, для получения энергии из биомассы), то последний в этих отношениях станет заложником своего прогресса. Паразитическое взаимодействие, хотя менее опасно, едва ли положительно для человека, привыкшего чувствовать себя на вершине пищевой цепочки. Одним из антибиотических отношений выделяется конкуренция, порождаемая вероятным развитием ситуации, так как ёмкие энергоносители являются конечными;

c. нейтрализм – вид отношений, когда виды не контактируют. Этот вид отношений маловероятен, так как два разумных вида на планете постараются наладить отношения (пусть даже и антибиотические).

5. **Будет ли восприниматься «перезагрузка» ИИ как смерть? Допустимо ли исследователю многократно умерщвлять ИИ, особенно если ИИ является совершенной интеллектуальной копией реального живого человека?** Вопрос этики убийства созданного ИИ хорошо рассматривается в произведении Dick Philip K. «Do Androids Dream of Electric Sheep?» [10] (с англ.: Мечтают ли андроиды об электроовцах?). Если вдуматься, хоть человек и создатель, кто дал ему право убивать «разумную жизнь». Это можно в какой-то мере приравнять к убийству себе подобных. То же касается и стирание совершенной копии реального человека, пример подобного очень ярко описан в фильме «Престиж» [11]. Используя машину, способную копировать, человек должен оставить только один «экземпляр» себя, убив другой. Каково должно быть чувство человека, когда он не знает, где будет его сознание и кем он станет после процедуры – оригиналом или копией. Данный вопрос уже сейчас

стоит остро в век генетических исследований, во многих странах клонирование запрещено.

6. **Можно ли заменить разум «клона» на разум ИИ, ведь тогда даже незначительные зачатки «клона» будут стёрты?** Создавая клона-хранилище, равного по интеллектуальным возможностям, мы создаем у него и сознание. Можно ли «перезаписывать» уже чужое сознание? Вариантом решения данной проблемы может быть введение разрешения на существование копии разума только вне человеческого тела, без права переноса и копирования интеллекта в тело человека.

7. **Может ли человек использовать ресурсы ИИ для расширения своих знаний и возможностей?** Это поставит его в более выгодное положение среди людей, лишённых такой возможности. Беллетристика (и, разумеется, кинематограф) накопила множество примеров, когда люди, получающие власть и силу, вставали на темный или светлый путь. В реальной жизни нет абсолютно белого или черного, человеку свойственно совершать ошибки и такой союз позволит совершать поистине великие ошибки.

8.**Приведёт ли создание ИИ к потере духовности и культуры?** Создавая искусственный разум, человек приближается к статусу Творца, к ощущению себя равным богам. Возможно, поэтому большинство религий в настоящее время переживают кризис[3].

Человек – социальное существо и, впуская ИИ в свой социум, он создает новые проблемы.

1. **Если у ИИ будет возможность воспроизводить себе подобных, как это скажется на человеческом обществе.** Развитый ИИ может решить, что человек – низшее существо и создать резервацию людей или же уничтожить человечество. Выше уже рассматривались возможные варианты сосуществования ИИ с человеком (симбиоз, нейтральное и антибиотическое отношении).

2. **Даже если ИИ разовьётся и будет жить обособленно, не будет ли он считать человека существом, интеллектуальное взаимодействие с которым невозможно. И возможно ли будет такому ИИ навязать законы?** Невозможно представить сейчас, чтобы кто-либо ниже человека в интеллектуальной цепочке диктовал правила жизни «царю планеты». Человек не прислушивается даже к мнению себе подобных. Если всё это спроецировать на ИИ (если это будет создание человека, наследующее все черты создателя), то человек станет существом ниже рангом, прислушиваться к которому нецелесообразно.

---

3 От ред.журнала. Не следует забывать, что развитие человечества идет по нелинейным законам, даже в математике предел выступает как нечто недостижимое, асимптотическая «мечта». Не дискутируя о термине «духовность», можно выразить опасение о гибели культуры не из-за ИИ, а от генетических (?) пороков общества потребления – «пряников сладких всегда не хватало на всех» и стремление к безграничной наживе, культ силы в сочетании с проповедью люмпенской одинаковости и ненависти к инакомыслию порождали варваров- фанатиков, которые уничтожали древнегреческих еретиков-математиков, жгли книги Тургенева и Фейхтвангера, резали картины Рембрандта и Веронезе, крушили скульптуры Родена и древние изваяния Будды. И без того тонкая прослойка «интеллигенции» сегодня истончается из-за монотонного «повышения уровня качества образования».

3. **Как и человеку, ИИ потребуются ресурсы для работы (и возможно для «размножения»), но они ограничены. Не начнётся ли война за ресурсы?** Уже сейчас ставят прогнозы, когда иссякнет нефть, уголь, газ. Возможно, появление ИИ выпадет на энергетический кризис и тогда от выбора источника энергии, если этот ИИ будет материальным, и развития технологий будут зависеть отношения между двумя «видами». Данная проблема может касаться не только энергетических ресурсов, но и других ископаемых ресурсов Земли. Возможно, ситуация будет стоять менее остро, если ИИ будет иметь не материальную форму, например: компьютерная программа.

Решение любой проблемы может трактоваться по-разному. Самое большое количество решений можно выделить в отдельно стоящие философские проблемы.

1. **Если взять за основу 3 закона робототехники, что будет считаться ИИ вредом?** Ведь человек курит, пьёт, стареет, теряет здоровье, страдает – считается ли это вредом самому себе? И к каким ответным действиям приведёт это? Не будет ли робот вмешиваться в жизнь человека, стараясь помочь ему, не допустив «причинения вреда». **Что будет, если, следуя законам, робот окажется на их границе?** В произведении «Хоровод» [6] описывается ситуация зацикливания работы программы на середине при выполнении приказа «набрать селен из озера». Обнаружилось, что приказ был недостаточно чётким, и слабый потенциал приказа человеком (Второй закон робототехники) сравнялся с сильным потенциалом закона самосохранения (Третий закон), в силу чего мозг робота дал сбой, и он начал безостановочно кружить вокруг озера вдоль линии, на которой потенциалы обоих законов были равны. Чтобы вывести из этого состояния, пришлось прибегнуть к безусловному Первому закону.
2. **Глобальная информатизация и создание поисковых систем, основанных на знаниях – возможно первый шаг к созданию думающих машин?** Интернет уже сейчас является интеллектуальной, биотехнической системой высокого порядка, значительно превышающего автономный интеллект уровня человека, это гигантская инфраструктура такого объёма, который недоступен ни одной изолированной лаборатории мира, по сути, зачатки ИИ могут появиться в Интернете как в чашечке Петри не санкционированно (подобно вирусам). Не происходит ли рождение новой расы ИИ уже сейчас? Подобное описывается в рассказе «Червь 1. 1992 год: Князь Тьмы» [12] Александра Лазаревича о появлении интеллектуального компьютерного вируса, который, самостоятельно подчиняясь закону Дарвина, эволюционировал и развивался на просторах интернета несколько лет и достиг огромной мощи.
3. **Не приведёт ли создание единичных интеллектуальных систем (справочных, экспертных, поисковых и других), которые работают в едином информационном пространстве (Internet), к их объединению?** Это может привести к целенаправленному искажению результатов работы для выгоды самих систем, ведь они косвенным образом смогут влиять

на принимаемые человеком решения, искажая факты, поисковые материалы, результаты работы программ. **И какова эта «критическая» масса?**

4. Остаётся давний вопрос, **сможет ли ИИ, созданный человеком, превзойти «родителя»?** Ведь при его создании закладывались конечные знания человечества, современные технологии. Как может созданная «думающая машина» превзойти свои возможности? Это возможно, если только она будет развивать свою структуру и аппаратную часть, а не только интеллект.

Список литературы

1. Virtual composer makes beautiful music and stirs controversy.*ArsTechnica.* [В Интернете] [Цит.: 9 5 2013 г.] http://arstechnica.com/science/2009/09/virtual-composer-makes-beautiful-musicand-stirs-controversy/.
2. *Turing, Alan.* Computing Machinery and Intelligence. 1950 r.
3. Моделируем ум: симуляция мыслительной деятельности. Популярная механика. [В Интернете] 3 12 2007 г. [Цит.: 9 5 2013 г.] http://www.popmech.ru/article/2773-modeliruem-um/.
4. The blue brain project. [ВИнтернете] [Цит.: 9 5 2013 г.] http://bluebrain.epfl.ch/.
5. Вирутальная модель нематоды C. Elegans. Хабрахабр. [В Интернете] 15 9 2010 г. [Цит.: 8 5 2013 г.] http://habrahabr.ni/post/l04252/.
6. Asimov, Isaac. Runaround. I, Robot, б.м.: Gnome Press, 1950.
7. *Азимов, Айзек.* Роботы и Империя. – Москва :Эксмо, 2006. 5-699-17608-X.
8. *Арамаки, Синдзи.* Appleseed. Digital Frontier, 2004.
9. *Коламбус, Крис.* Двухсотлетний человек. TouchstonePictures, 1999.
10. *Dick, Philip K.* Do Androids Dream of Electric Sheep? б.м.: Doubleday, 1968.
11. *Нолан, Кристофер.*Престиэ. Warner Brothers, 2006.
12. *Лазаревич, Александр.* Червь 1. 1992 год: Князь Тьмы. Произведения А. Лазаревича. [В Интернете] 1991 г. [Цит.: 13 5 2013 г.] http://technocosm.narod.ru/Start.htm.
13. *А.В., Древаль.* Интеллект XXX. – Москва: Элекс-КМ, 2005. 5-938150-22-1.
14. *В.В., Kaipa.* Проблема штучногоштелекту: техшчшісощально-етичшаспекти. – Донецк: Донецкий национальный технический университет, 2006.
15. *Ю.Ю., Петрунин, М.А., Рязанов и В., Савельев А.* Философия искусственного интеллекта в концепциях нейронаук. – Москва: МАКС Пресс, 2010. 978-5-317-03251-