УДК 141.3

Сознание и искусственный интеллект: гипотезы и прогнозы

Ефимова Ирина Яковлевна

Кандидат философских наук,
доцент кафедры гуманитарных и естественно-научных дисциплин,
Московский институт психологии,
129223, Российская Федерация, Москва, просп. Мира, 119, стр. 501;
e-mail: inef_ina@mail.ru

Аннотация

Статья посвящена аналитическому обзору современных теорий искусственного интеллекта, включая четыре оригинальных концепции XX-XXI вв., вызвавшие наибольший резонанс в научном сообществе. Исследуя эволюцию компьютерных технологий, автор подчеркивает, что первоначально данные технологии создавались для исследования процессов деятельности мозга человека, поэтому закономерно возник вопрос о возможности научить машину мыслить. Эта проблема разделила ученых на два непримиримых лагеря, с одинаковой убежденностью отстаивающих два противоположных тезиса. В споре «за» и «против» до сих пор не найден консенсус, вследствие чего вопрос остается открытым. В число современных перспективных изысканий входит эволюционная кибернетика, исследующая эволюционное происхождение интеллекта. Открытия, сделанные в данной области, внесли значительный вклад в развитие нанотехнологий, робототехники и других отраслей. В. Турчин, рассматривая когнитивную и биологическую эволюцию, пропагандирует системный подход в устройстве мира, главным принципом которого является метасистемный переход. Другой концепцией является цифровая философия Э. Фредкина, провозгласившего дискретность всех процессов в природе, что позволяет свести все к обработке информации и последующему компьютерному моделированию. Эту гипотезу дополняет концепция аргумента моделирования Н. Бострома, который утверждает, что современная цивилизация сама является продуктом моделирования, следовательно, вполне вероятно, что наш мир следует рассматривать как матрицу. Завершает обзор футурологический прогноз А. Нариньяни, предрекающего в ближайшей перспективе кибернетизацию человечества и превращение его в новый вид eHomo как симбиоз компьютера и человека.

Для цитирования в научных исследованиях

Ефимова И.Я. Сознание и искусственный интеллект: гипотезы и прогнозы // Контекст и рефлексия: философия о мире и человеке. 2017. Том 6. № 6А. С. 235-246.

Ключевые слова

Искусственный интеллект, интенциональность, эволюционная кибернетика, нанотехнология, цифровая философия, аргумент моделирования, матрица.

Введение

Понятие искусственного интеллекта (ИИ) было введено в 1956 г. американским ученым Дж. Маккарти и обозначало совершенную техническую систему, моделирующую функции человеческого мышления, при этом моделирование понималось не как воспроизведение мыслительных функций, а как их имитация. Создание ИИ первоначально предназначалось для исследования процессов, происходящих в мозге человека, отсюда возникли такие комплексные направления, как психофизиология, нейробиология, биоинформатика, эволюционная кибернетика и многое другое. Одним из самых острых вопросов, касающихся проблемы ИИ, является вопрос о том, можно ли научить машину мыслить. В дискуссиях, развернувшихся вокруг этого, часть исследователей утверждает, что последнее невозможно, поскольку (и здесь идут в ход такие атрибуты сознания, которые ранее объявлялись второстепенными) машина не имеет таких качеств, как интуиция и бессознательное. В известном аргументе «китайской комнаты» Серла [Серл, 2002, 61] утверждается, что ИИ ни при каких обстоятельствах не поднимется до уровня человеческого сознания по причине функциональной неспособности машины сократить разрыв между изначально заданной и производной интенциональностями. ИИ (при соответствующем программном обеспечении) может создать иллюзию мыслительной активности, значительно превышающей способности человеческого сознания, но ИИ лишен творческого потенциала мышления, присущего человеку, постоянно генерирующему новые (а часто и неожиданные) смысловые значения. Противоположная точка зрения, которой придерживается Д. Деннетт, напротив, заключается в том, что ИИ считается равным человеческому сознанию по принципу подобия (т. е. они оба не обладают первичной интенциональностью). Таким образом, сходство с человеческим мышлением присутствует не в ИИ, а, напротив, в человеке, мыслящем подобно ИИ [Dennett, 1987]. Рассматривая возникновение интенциональных содержаний в сознании человека на примере установленных правил манипуляций с объектами, Деннетт приходит к выводу о том, что ИИ ни в чем не уступает человеку. Возможно, в будущем программы нового поколения в системе ИИ смогут настолько интегрироваться в человеческую жизнедеятельность, что при обоюдном соблюдении установок и правил, разграничивающих функции, ИИ в буквальном смысле научится осмысливать, понимать и давать оценку происходящим вокруг него событиям.

В дальнейшем сложилось достаточное количество гипотез о возможностях искусственного интеллекта, которые также сводятся к двум противоположным тезисам: ИИ равен сознанию (Д. Деннетт, Д. Хофштадтер, Р. Джекедофф и т. д) и ИИ не равен сознанию (Дж. Серл, Р. Пенроуз, Х. Дрейфус, Г. Хант и т. д.). В целом в науке не выработалось единого мнения по поводу данного вопроса хотя бы потому, что для определения интеллекта искусственного необходимо иметь четкое представление об интеллекте естественном, о котором современная наука также имеет весьма приблизительное понятие.

Эволюционная кибернетика и исследование искусственного интеллекта

Один из подходов в исследовании ИИ связан с направлением эволюционной кибернетики (синергетики), которая рассматривает эволюцию систем управления и кибернетических организмов (т. е. фактически предпринимается попытка обосновать эволюционное происхождение интеллекта). Здесь следует сказать о двух развивающихся (возникли в конце 1980-1990 гг.), взаимосвязанных направлениях кибернетических исследований: «Искусственная жизнь» (Artificial Life) и «Адаптивное поведение» (Adaptive Behavior) [Редько, 2005, 150-151].

Главным стимулом к проведению подобных изысканий послужило стремление осмыслить принципы структурирования органической жизни, в целях чего и была создана соответствующая модель. По утверждению одного из представителей этого направления К. Лангтона, гипотеза существования искусственной жизни построена на предположении, что материальная и содержательная (логическая) формы могут быть отделены друг от друга [Langton, 1992, 41]. Приверженцы линии «Искусственная жизнь» убеждены, что в их опытах наглядно демонстрируются универсальные принципы консолидации организмов безотносительно к земным условиям. Это значит, что объектом исследования является вероятностная, а не эмпирически доступная форма бытия. В качестве образца наиболее характерных исследований искусственной жизни можно привести следующие примеры:

1) модель Полимир (Poly World) Л. Ягера представляет собой компьютерную версию искусственной жизни, имеющую ряд характерных особенностей, в частности организованную нейронную связь, цветовое зрение, движение, умение находить пищу для пополнения энергетических ресурсов, способность к образованию симбиозов и противоборств. В ходе наблюдений за общей эволюцией также имели место эпизоды неординарной поведенческой реакции;
2) модель Тьерра (Tierra) Т. Рэя – модель эволюции компьютерных организмов, способных к самовоспроизведению. Объекты программы имеют внутри себя геном, отвечающий за выполнение определенных задач и способствующий последующему видовому усложнению системы;
3) модель Авида (Avida) была создана программистом К. Адами на основе Тьерры. С ее помощью были получены характеристики распределения особей в эволюционирующих популяциях, на основе чего был выявлен факт скачкообразного, а не непрерывного развития эволюции;
4) модель Эхо (Echo) – модель, созданная проектировщиком Дж. Холандом, которая отображает процесс эволюции примитивных систем, взаимодействующих друг с другом по схеме элементарного социума: торговля, война, скрещивание особей, которые приводят, однако, к возникновению более сложных социальных феноменов, например, в виде «мировых войн» или тотальных объединений [Редько, 2005, 151].

В начале 1990-х гг. начинает активно развиваться исследовательская программа «Адаптивное поведение» (АП), основными задачами которой являлись моделирование и изучение искусственно созданных объектов (например, роботов или компьютерных программ), демонстрирующих способность адаптации по отношению к окружающему миру. Данные устройства были названы «аниматами» по аналогии с настоящими животными, поведение которых должен имитировать анимат искусственный. Исследовательская программа-максимум «Адаптивного поведения» ставит пред собой цель проследить последовательность развития ментальных данных у животных и проанализировать природу эволюции человеческого сознания. Для обоих направлений («Искусственная жизнь» и «Адаптивное поведение») свойственен синтетический подход в решении ряда задач, общим для них является и использование нетривиальных компьютерных методов: нейронных сетей, генетического алгоритма, классифицирующих систем и т. д. [Там же, 150-151].

Рассмотрим далее наиболее выдающиеся концепции XX-XXI вв., связанные с теорией эволюции и исследованиями в сфере ИИ. Следует обратить внимание на то, что, несмотря на полную независимость друг от друга, все представленные гипотезы имеют смысловую взаимосвязь.

Теория метасистемных переходов

В. Турчин – создатель эволюционной кибернетики и автор теории метасистемных переходов [Турчин, 2000, 350-356]. В книге «Феномен науки. Кибернетический подход к эволюции» Турчин анализирует процесс эволюции двух диаметрально противоположных систем: биологической и когнитивной, при этом первая рассматривается с позиции кибернетики, а вторая – как закономерное следствие первой. Таким образом, пропагандируется иерархический системный подход к устройству мира, согласно которому каждый объект является системой, состоящей из взаимосвязанных и работающих в совокупности подсистем. Главный феномен в системной организации мира – принцип метасистемного перехода. В «Кибернетическом манифесте», прилагающемся к книге, Турчин дает следующую характеристику ключевых идей своей работы [Там же].

Метасистемный переход возникает в том случае, когда определенное количество независимых систем объединяются в одну организацию с принципиально новым типом управления. Вновь образованная система поднимается до уровня метасистемы в сравнении с прежней. Таким образом, совершается активированный извне процесс творческой эволюции (поэтому Турчин называет акт метасистемного перехода «квантом эволюции»). Вследствие ряда переходов различного уровня все метасистемы претерпевают ряд иерархических преобразований как внутри собственной структуры, так и в системе внешнего управления. Следовательно, самые масштабные скачки эволюции произошли в результате глобальных метасистемных переходов в процессе отсеивания природой всего лишнего. По этому принципу (усложнения) возникло человеческое сознание со всем набором функций, отличающих его от животного, и, наконец, кульминацией процесса эволюции явились зарождение и развитие цивилизации. Турчин убежден, что человечество преодолело природный механизм естественного отбора в биологической эволюции и способно творчески «направлять» и «исправлять» собственную эволюцию посредством сознательного усилия, опираясь на глубинное понимание процессов, происходящих в ней.

Дальнейший процесс социальной интеграции неизбежно приведет цивилизацию к слиянию культур и формированию неделимого государства в масштабах Земли, а также к созданию централизованного планетарного правительства, осуществляющего управление и контроль за использование природных ресурсов. Интеграция затронет не только социальную сферу, но и непосредственно людей, тела которых будут интегрированы и генетически запрограммированы на неопределенно долгое время, в потенциале – на бесконечность. Рассматривая человека как сложную кибернетическую систему, которой управляет сознание, Турчин выдвигает предположение о том, что в будущем для обеспечения фактического бессмертия сознание можно будет перемещать в другие носители искусственного или биологического происхождения. Первостепенная задача – сохранение личностной индивидуальности, бесконечное расширение границ ее творческого потенциала в целях дальнейшего развития всей метасистемы человечества в планетарном значении. С точки зрения эволюционной целесообразности прерывание жизненного цикла необходимо только для простых организмов, тогда как человек – вершина эволюции – не должен ограничивать свое развитие рамками одной жизни (в среднестатистическом значении). Именно поэтому основная цель эволюции – бессмертие каждой человеческой личности. В завершение манифеста ученый повторяет, что наша цивилизация стоит на пороге метасистемного перехода, этот факт указывает на эволюционную необходимость скорейшей консолидации человечества при сохранении творческой независимости индивида. Если этого не произойдет, неизбежен процесс деградации и вымирания человечества как вида.

Цифровая философия

Следующей концепцией, основанной на базе открытий в области ИИ, является цифровая философия (digital philosophy) Э. Фредкина, опубликовавшего в 1992 г. две работы, опровергающие прежние представления о Вселенной, согласно которым все процессы во времени и пространстве являются непрерывными и текучими [Fredkin, 1992; Fredkin, 1993; Fredkin, 2003]. Основным тезисом гипотезы Фредкина является дискретность всех процессов в природе (т. е. дискретны не только все материальные вещи, состоящие из отдельных молекул и атомов, но и энергия электричества и света, которая также состоит из разрозненных «частей», или квантов). По мнению Фредкина, этот факт позволяет теоретически свести все физические и ментальные процессы (включая сознание) к обработке информации и их последующему компьютерному моделированию [Rhodes, 2000, 6-35]. Когда Фредкин употребляет термин «информация», он подразумевает «скалярное количество», т. е. нечто такое, что может быть представлено точкой на шкале. Свойства, которые могут быть определены количественно, можно выразить в цифрах. Таким образом, под ограниченностью природы Фредкин предполагает, что все ее свойства могут быть полностью выражены цифрами, поскольку эти свойства носят дискретный и поэтапный характер. При этом Фредкин рассматривает полученные данные как соответствующую информацию, а не как смысл, связанный с данными. Отсюда любая информация может быть легко выражена числами как символами слов или предложений (соответствующими 0 и 1). Интерпретации этих цифр легко добиться, внимательно отслеживая, какому символу соответствует каждая цифра. Фредкин приводит следующий пример: три одинаковых монеты могут быть представлены в форме различных символов. Ничего не добавилось бы к этой информации при использовании реальных монет, таким образом можно сэкономить много металла, заменив его на символы. Учитывая ограниченность природы, в буквальном смысле не существует никакой разницы между информационным содержанием объема пространства-времени и «вещью», потому что информация есть вещь и наоборот.

Фредкин отмечает, что один особый вид компьютерной архитектуры представляет собой полную аналогию концепции дискретных элементов космического времени, которое развивается как индивидуально, так и коллективно от одного момента до другого. Эта архитектура обозначает каждую вычислительную единицу как «ячейку», которая работает, как если бы это был независимый компьютер с простым набором программных инструкций. Поскольку она работает независимо и автоматически, подобно роботу со своим набором правил, ячейка обозначается как «автомат». Скопление многих из этих вычислительных ячеек и составляет «Клеточные Автоматы» компьютерной архитектуры. Изучение компьютерной архитектуры показывает, что она обладает огромным потенциалом для моделей всех видов взаимодействия в природном мире. Подобно клеточному автомату, квантовая единица в нашем мире взаимодействует с подобными ей единицами, производя изменения согласно строгим математическим правилам, учитывая не только собственное дискретное, индивидуальное состояние, но и состояние окружающих ее единиц. В итоге Фредкин заявляет: «Учитывая ограниченность (дискретность) природы, мы, по сути, являемся своего рода клеточными автоматами. И так как Автоматы, Клеточные Автоматы и другие машины – все формы компьютеров, это означает, что в основе физики естественного мира мы имеем компьютер в том или ином виде… поскольку то, что не может быть запрограммировано, не является физикой» [Fredkin, 2003, 190].

Цифровую философию Э. Фредкина часто называют современной интерпретацией монистической метафизики Г. Лейбница с той лишь разницей, что монады заменяются теорией клеточных автоматов. В цифровой Вселенной существование и мысли – все становится эквивалентным вычислению, включая сознание, которое (как простую информационную единицу) также можно подвергнуть вычислительной обработке.

Концепция аргумента моделирования

Гипотезой, косвенно подтверждающей выводы Фредкина, является концепция аргумента моделирования (simulation argument) британского философа Н. Бострома из Оксфорда, который утверждает с вероятностью три к одному, что (коль скоро сознание можно подвергнуть компьютерному моделированию) современная цивилизация сама может оказаться продуктом моделирования. В одной из последних статей «Мы живем в компьютерном моделировании?», изданной в академическом научном журнале «New Scientist», Н. Бостром выдвигает логическое обоснование своей теории [Bostrom, 2006].

В качестве предварительного замечания к центральному тезису Бостром отмечает, что многочисленные открытия науки о мире и человеке не являются поводом для оптимизма. Земля – это не центр вселенной, вид Homo sapiens произошел от животных, человек пронизан нейрофизиологическими импульсами и подвержен разнообразным биологическим, психологическим и социологическим влияниям, над которыми имеет ограниченный контроль и понимание. Технологический прогресс, который был достигнут благодаря прежним поколениям, также в некотором смысле уничижителен, так как предполагает, что наиболее развитая технология, которая существует сегодня, чрезвычайно ограничена и примитивна по сравнению с той, что будет иметь место в будущем. Если экстраполировать этот ожидаемый технологический скачок и продумать некоторые из его логических последствий, можно прийти к другому неутешительному прогнозу, который Бостром обозначил как «аргумент моделирования» (simulation argument). Формальная версия аргумента нуждается в некоторой теории вероятности, но основная идея может быть схвачена без математики. Она начинается с предположения о том, что будущие цивилизации могут располагать настолько развитой компьютерной технологией и возможностями программирования, что будут способны создать «модель предка» (ancestor simulations) или так называемую матрицу (matrices) [Bostrom, 2003], т. е. они смогут детально смоделировать копии предшественников (предков), воссозданных настолько подробно, чтобы сознавать и иметь те же самые виды опыта, которые имеем мы.

Зачем понадобится создание Матрицы? Возможно, предполагает Бостром, будущие историки создали бы Матрицу, которая подражала бы разновидности их собственной истории, и могли бы больше узнать об их прошлом и исследовать противоречащие фактам исторические сценарии. Возможно, матрица создается как художественный сценарий (подобно многосерийному фильму) или туристическая промышленность будущего моделирует наиболее интересные исторические эпохи, для того чтобы их современники могли погрузиться в прошлое, вступая в моделирование и взаимодействуя с его жителями. Возможные поводы являются бесчисленными. Если будущих людей представить, как существующих, имеющих соответствующую технологическую базу и юридические права для создания Матрицы, то можно было бы ожидать, что огромное количество матриц было бы создано, включая ту, которая будет похожа на мир, в котором мы живем. Таким образом, Бостром предлагает представить матрицу как весьма реальную виртуальную возможность окружающей среды, где умы, населяющие мир, являются самостоятельной частью моделирования. В подобном

моделировании не все должно быть совершенным, но только достаточно приемлемым для его обитателей. Не было бы необходимости моделировать каждый объект до субатомного уровня: если предмет (например, книга) был бы смоделирован, то, согласно программе, в него достаточно было бы включить визуальное появление, вес, структуру и несколько других макроскопических свойств, потому что, используя данный предмет, мы не интересуемся тем, как расположены его индивидуальные атомы в этот момент времени. Если бы потребовалось изучить предмет более тщательно, допустим под микроскопом, то дополнительные детали моделирования могли быть воссозданы так, как это необходимо. Также можно представить, что только немногие люди моделируются достаточно подробно, для того чтобы быть сознательными, в то время как другие представляют собой подобие «полуфабриката». Они могут появляться и вести себя как реальные люди, но в действительности они являются всего лишь «зомби», функционирующей социальной единицей, которая не является способной к какой-либо рефлексии.

В заключительной части аргумента моделирования Бостром предлагает рассмотреть три основных тезиса в соответствии с правилами логики:

1) почти все цивилизации, достигнув нашего уровня развития, вымирали перед становлением технологической зрелости;
2) процент технологически развитых цивилизаций, которые интересуются созданием матрицы, почти равен нолю;
3) вы почти наверняка живете в компьютерном моделировании.

Чем обоснован последний вывод? Предположим, размышляет Бостром, первое суждение ложно. Тогда существенная доля цивилизаций, достигших нашего уровня развития, в конечном счете должна достичь технологической зрелости. Предположим также, что второе суждение является ложным. Тогда большая часть этих цивилизаций управляет моделями предков. Таким образом, согласно выводу Бострома, если первое и второе ложно, должны существовать моделируемые умы, которые подобны нашим (т. е., согласно логике, непоследовательно отклонять все три суждения, одно из них должно быть верным). Приведя в действие числа, можно обнаружить, что моделируемых умов должно быть значительно больше, чем оригиналов. Бостром выдвигает предположение о том, что технологически развившиеся цивилизации должны иметь доступ к столь огромным вычислительным ресурсам, что фактически, посвятив даже самую незначительную часть этих ресурсов моделированию матрицы, они были бы способны воссоздать миллиарды моделей всех когда-либо существовавших людей. Другими словами, почти все умы, подобные нашим, моделировались бы. Поэтому, рассуждая беспристрастно, каждый может предположить, что его личность, вероятно, представляет собой не оригинал, а одну из этих смоделированных копий.

Даже единственный компьютер (планетарного размера), построенный на основе передовой молекулярной технологии, сможет смоделировать всю интеллектуальную историю человечества при использовании меньше чем миллионной из ее вычислительной мощи в течение одной секунды, и этот факт выводится только на основе известных вычислительных машин. Отдельная же развитая цивилизация может строить миллионы таких компьютеров.

Рассматривая варианты более детально, Бостром рассуждает следующим образом. Суждение первое является прямым. Например, возможно представить некоторые технологии, до уровня которых каждая продвинутая цивилизация в конечном счете развивается и которые затем уничтожают ее. Суждение два указывает на то, что существует некое незыблемое соглашение (конвергенция) среди всех развитых цивилизаций, согласно которому ни одна из них не интересуется управляемым моделированием. Можно вообразить различные причины,

которые могут иметь цивилизации, чтобы сделать этот выбор. Однако данный факт свидетельствовал бы об известном ограничении будущей эволюции интеллекта. Третья возможность философски наиболее интригующая. Если она верна, мы почти наверняка являемся частью компьютерного моделирования, созданного некой продвинутой цивилизацией.

Таким образом, Коперник, Дарвин, Эйнштейн и другие известные ученые открыли законы, действующие в смоделированной реальности. Возможно, эти законы идентичны тем, которые работают на более фундаментальном уровне действительности, где существует компьютер, управляющий нашим моделированием (который, конечно, сам может быть моделью), а возможно – нет. Исходя из сказанного, наше место в мире может оказаться еще более скромным, чем предполагалось. Можно ли обнаружить «создателей» моделируемого мира? По мнению Бострома, и да, и нет. Если «симуляторы» хотели бы обнаружить себя, они бы это сделали (например, в Вашем персональном компьютере всплывет окно с текстом: «Вы живете в матрице, для получения необходимой информации щелкните здесь»); если они этого не хотят, то их, вероятно, никогда не обнаружат. Доказательство, которое позволит с высокой степенью вероятности заключить, что мы находимся в моделировании, появится в том случае, если человечество достигнет того уровня развития, в котором начнет моделировать собственные модели предков.

eHomo – два в одном

Не менее интересную гипотезу о будущем развитии человеческого интеллекта выдвигает профессор А. Нариньяни (директор НИИ Искусственного интеллекта, г. Москва). В докладе «eHomo – два в одном (Homo sapience в ближайшей перспективе)» ученый утверждает, что очень скоро (через 10-15 лет) сегодняшний цивилизованный Homo превратится в eHomo – новый вид, сохраняющий биологическую принадлежность к Homo sapience, но качественно значительно отличающийся от него за счет симбиоза с продуктами стремительно развивающихся сверхвысоких технологий [Нариньяни, 2005]. По оценке Нариньяни, в настоящее время темп развития компьютеров нарастает в соответствии с законом Мура [Moore, 1965], согласно которому скорость процессора удваивается каждые два года. Также быстро совершенствуются и другие составляющие компьютера – от объема памяти всех уровней до степени развития интерфейсов и периферии. Подобная прогрессия ИТ неизбежно должна повлиять на будущее цивилизации, сформировав принципиально новый вид человека (eHomo). Основными составляющими в формировании завтрашнего человека и его цивилизации, по мнению Нариньяни, будут являться следующие факторы: естественный отбор в мире электроники; роботы, внедренные в тело; симбиоз с потомком мобильного телефона; среда обитания eHomo; е-возможности формирования личности; глобализация индивидуального общения; eHomo и большой брат; eHomo крупным планом.

Каждой из представленных характеристик Нариньяни дает свою интерпретацию. Понятие «естественный е-отбор» связано с интеллектуализацией ИТ, данный факт обусловлен гиперактивностью двух тесно взаимосвязанных процессов. Первый процесс – это растущий поток различной е-техники, которая, множась и эволюционируя, борется за свое право на существование (т. е. место на рынке), пытаясь доказать свою полезность и необходимость. Второй процесс связан со всеобщей информатизацией, проникающей не только в бытовую жизнь, но и в тело человека. Для того чтобы не исчезнуть в этом потоке, различные технологии и другие автономные устройства от микро до макро должны сохранить себя в жестком естественном отборе, который уже исчерпал возможности периода вегетативного развития и

требует не только удешевления и миниатюризации, но и все более высокого уровня интеллекта. Информационные технологии вступают в эпоху, когда требуется мгновенное «понимание» пользователя, часто лучшее, чем понимает он сам.

На сегодняшний день уже имеются роботы (имплантаты), внедренные в тело. Высокие технологии, созданные для внедрения в организм, делятся на три уровня: долговременные или пожизненные составляющие, элементы длительного (недели, месяцы) и оперативного (часы, дни) действия. Данные технологии не предполагают прямого внедрения личности в некую суперсеть в стиле «Матрица». Однако, утверждает Нариньяни, в самое ближайшее время начнется все более широкое внедрение в организм датчиков и эффекторов, имеющих название нано- и даже молекулярных роботов.

Микроустройства типа чипов неизбежно будут вытеснены гораздо более тонкими элементными базами, приближенными к молекулярному и атомному уровням, известных как нано, био, гено. Эти технологии будут работать для реализации внутренних и внешних функций. Первые будут направлены на совершенствование «системы тела», что предполагает оптимизацию заданного природой организма и его «реинженеринг». Развитие внешних функций будет направлено на прямое взаимодействие организма eHomo со средой (например, на управление компонентов среды сигналами мозга). Другая элементная база, направленная к мозгу извне, обеспечит расширение возможностей личности и окажет воздействие на центральную нервную систему в лечебных целях, а также может использоваться для коррекции психики eHomo: ограничение агрессии, блокирование боли, мобилизация и т. д. (последнее имеет некий негативный аспект, так как открывает возможность для массового манипулирования).

Грядущий симбиоз с потомком мобильного телефона проявится в результате форсированного роста интеллектуализации ИИ. В итоге этого симбиоза к середине века eHomo от рождения до старости будет находиться в индивидуальном информационном коконе, выступающем в роли Alter ego личности и способствующем ее развитию и ментальному росту. Все более широкий канал информации, получаемой человеком извне, помноженный на кибернизацию организма и колоссальный прогресс виртуальной реальности, начнет размывать и без того нечеткую грань между объективным восприятием и субъективным, синтезируемым искусственно (включая тактильные ощущения, запахи и эмоции).

Наступление Новой эры будет иметь и негативные последствия, ограничивая свободу eHomo, поскольку чем выше технический уровень Ноосферы, тем более зависимым от ее многомерной сложности становится eHomo. Его индивидуальность контролируется не только отпечатками пальцев, но и различными биометрическими устройствами, идентифицирующими голос, зрачок, ДНК. Возможно впоследствии внедренный в тело чип-идентификатор будет не только сообщать личный набор параметров и хромосом, но и контролировать передвижения в пространстве с точностью до метра. Таким образом, «плохому» или асоциальному eHomo будет все сложнее уходить от контроля «Большого брата», поскольку он потеряет возможность сменить личность или «исчезнуть из обращения». При этом мера определения хорошего и плохого будет устанавливаться с точки зрения Новой этики надвигающейся е-цивилизации. В новом мире уйдут из обращения деньги, перейдя в категорию виртуальную, отсюда личные финансы и имущество, а также вся экономика, основанная на рынке, станут абсолютно прозрачны, превратившись в подобие одинаковых (хотя и конкурирующих) компьютерных программ.

Нариньяни утверждает далее, что новые технологии, по сути, открывают путь к индивидуальному бессмертию. Этому сопутствуют два фактора: быстро расширяющийся объем

и разнообразие оцифрованной «персональной» информации и доступность неограниченных ресурсов памяти для ее накопления. Информация становится неотделимой частью среды обитания. В настоящее время разрабатываются технологии, превращающие обычный мобильный телефон в устройство, записывающее жизнь человека, автоматически систематизируя все проходящие через личные информационные каналы сообщения, снимки, видео и звуковые клипы. Сюда же планируется внедрить устройство для воспроизведения трехмерного пространства, запахов, ощущений, а также считывания информации с электронных элементов, контролирующих общее состояние организма. Если к этому добавится возможность визуализировать и фиксировать внутренний виртуальный мир – сны, воспоминания, воображаемые образы, то полнота и воспроизводимость этой информации в буквальном смысле отделит душу от тела. При этом именно ментальная составляющая (или душа) получит в активной е-памяти 0,99… alter ego, а тело превратиться в сложное устройство для записи и воспроизведения этого мегаархива.

В число прочих сопутствующих технологий следует включить клонирование и генную инженерию, которые позволят обновлять тело по частям или целиком, устраняя наследственные недостатки с полной гарантией бессмертия души на протяжении всего периода цивилизации, способной обеспечить стабильность такого процесса. Уже сегодня на будущее бессмертие работают быстро развивающиеся нанотехнологии, ориентированные на создание устройств молекулярного размера, способных следить за функционированием клеток и осуществлять их коррекцию в случае отклонений.

Свою гипотезу будущей эволюции человеческого и искусственного интеллектов профессор Нариньяни предлагает рассматривать как достаточно сходную с грядущим оригиналом, подчеркивая, что приближающаяся точка цивилизационной бифуркации (всегда отличающаяся своей внезапностью) eHomo слишком близка, чтобы относить ее к сфере научной фантастики.

Повторим, что четыре широко представленные выше концепции (В. Турчин, Э. Фредкин, Н. Бостром, А. Нариньяни) отражают, несмотря на их нестандартность, точку зрения академических ученых. Таким образом, данные гипотезы носят абсолютно легитимный характер и имеют широкий резонанс в научном мире. По мнению Б. Гейтса (создатель компании Microsoft), все четыре сценария могут быть реализованы (по крайней мере, один из них), а значит, в ближайшем будущем человечество ждут существенные перемены.

Заключение

Без сомнения, пока человек способен реализовать сознательный выбор, альтернатива представленным сценариям существует. Доказательством тому служит хотя бы то обстоятельство, что сверхспособности изначально заложены в сознание человека, но в силу разных причин у большинства людей они находятся в спящем состоянии. Эти иррациональные силы пробуждаются по мере сознательного усилия, включающего духовное и интеллектуальное развитие личности. Цивилизация, выбравшая подобный путь эволюции, могла бы превзойти все мыслимые перспективы техногенной эпохи. Ярким свидетельством этого может послужить всем известный факт необычайного расцвета теоретического мышления у древних греков при относительно низком развитии техники. Греки явились родоначальниками теоретического мышления и научного знания, ориентированного на самопознание и поиск истины ради самой истины, а не ради достижения каких-либо прикладных результатов, служащих целям технического прогресса, но уводящих человека от потребностей души.

Библиография

1. Нариньяни А.С. eHomo – два в одном (Homo sapience в ближайшей перспективе) // Материалы Всероссийской междисциплинарной конференции «Философия искусственного интеллекта». М., 2005. С. 378-392.
2. Редько В.Г. Эволюция. Нейронные сети. Интеллект. М.: УРСС, 2005. 220 с.
3. Серл Дж. Открывая сознание заново. М.: Идея-Пресс, 2002. 256 с.
4. Турчин В.Ф. Феномен науки. Кибернетический подход к эволюции. М.: ЭТС, 2000. 368 с.
5. Bostrom N. Are you living in a computer simulation? // Philosophical quarterly. 2003. Vol. 53. No. 211. P. 243-255.
6. Bostrom N. Do we live in a computer simulation? // New scientist. 2006. No. 3 P. 8-9.
7. Dennett D. The intentional stance. Cambridge, 1987.
8. Fredkin E. A new cosmogony // Proceedings of Workshop on physics of computation. New York, 1993. P. 116-121.
9. Fredkin E. An introduction to digital philosophy // International journal of theoretical physics. 2003. Vol. 42. No. 2. P. 189-247.
10. Fredkin E. Finite nature // Proceedings of the 27th Recontre de Moriond. France, 1992.
11. Langton C.G. Life at the edge of chaos // Langton C.G. et al. (eds.) Articial life II. Redwood City: Addison Wesley, 1992. P. 41-91.
12. Moore G. Cramming more components onto integrated circuits // Electronics. 1965. Vol. 38. No. 8. P. 114-117.
13. Rhodes S. The finite nature hypothesis of Edward Fredkin. Canton, 2000.
14. Turing A. Computing machinery and intelligence // Mind. 1950. No. 59. P. 433-460.

Consciousness and artificial intelligence: hypotheses and predictions

Irina Ya. Efimova

PhD in Philosophy,
Associate Professor at the Department of humanities and natural sciences,
Moscow Institute of Psychology,
121170, p. 501, 119 Mira av., Moscow, Russia Federation;
e-mail: inef_ina@mail.ru

Abstract

The article aims to carry out an analytical review of modern theories of artificial intelligence, including four original concepts developed in the 20th and 21st centuries, which caused the greatest resonance in the scientific community. Exploring the evolution of computer technology, the author emphasises that, initially, this technology was created to study processes of human brain activity, so there arose the question about the possibility to teach a machine to think, and the issue remains controversial. Evolutionary cybernetics, exploring the evolutionary origin of intelligence, is considered to be one of the most promising research areas. Discoveries in this field have made a significant contribution to the development of nanotechnology, robotics and other industries. V. Turchin promotes a systemic approach to the world, the main principle of which is a metasystem transition. There is another concept called digital philosophy. It was developed by E. Fredkin that pointed out the discreteness of all processes in nature, which leads to reducing everything to information processing and subsequent computer modelling. This hypothesis is complemented by the idea of the simulation argument (N. Bostrom) that views modern civilisation as a product of simulation. A. Narinyani in his futurological forecast predicts the cybernetisation of humanity in the short term and its transformation into a new kind of eHomo as man-computer symbiosis.

For citation

Efimova I.Ya. (2017) Soznanie i iskusstvennyi intellekt: gipotezy i prognozy [Consciousness and artificial intelligence: hypotheses and predictions]. *Kontekst i refleksiya: filosofiya o mire i cheloveke* [Context and Reflection: Philosophy of the World and Human Being], 6 (6A), pp. 235-246.

Keywords

Artificial intelligence, intentionality, evolutionary cybernetics, nanotechnology, digital philosophy, modelling argument, matrix.

References

1. Bostrom N. (2003) Are you living in a computer simulation? *Philosophical quarterly*, 53 (211), pp. 243-255.
2. Bostrom N. (2006) Do we live in a computer simulation? *New scientist*, 3, pp. 8-9.
3. Dennett D. (1987) *The intentional stance*. Cambridge.
4. Fredkin E. (1993) A new cosmogony. *Proceedings of Workshop on physics of computation*. New York, pp. 116-121.
5. Fredkin E. (2003) An introduction to digital philosophy. *International journal of theoretical physics*, 42 (2), pp. 189-247.
6. Fredkin E. (1992) Finite nature. *Proceedings of the 27th Recontre de Moriond*. France.
7. Langton C.G. (1992) Life at the edge of chaos. In: Langton C.G. et al. (eds.) *Articial life II*. Redwood City: Addison Wesley, pp. 41-91.
8. Moore G. (1965) Cramming more components onto integrated circuits. *Electronics*, 38 (8), pp. 114-117.
9. Narin'yani A.S. (2005) eHomo – dva v odnom (Homo sapience v blizhaishei perspektive) [eHomo – two in one (Homo sapience in the near future)]. *Materialy Vserossiiskoi mezhdistsiplinarnoi konferentsii "Filosofiya iskusstvennogo intellekta"* [Proc. Conf. "The philosophy of artificial intelligence"]. Moscow, pp. 378-392.
10. Red'ko V.G. (2005) *Evolyutsiya. Neironnye seti. Intellekt* [Evolution. Neural networks. Intelligence]. Moscow: URSS Publ.
11. Rhodes S. (2000) *The finite nature hypothesis of Edward Fredkin*. Canton.
12. Searle J. (1992) *The rediscovery of the mind*. Cambridge, MA: The MIT Press. (Russ. ed.: Searle J. (2002) *Otkryvaya soznanie zanovo*. Moscow: Ideya-Press Publ.)
13. Turchin V.F. (2000) *Fenomen nauki. Kiberneticheskii podkhod k evolyutsii* [The phenomenon of science. A cybernetic approach to evolution]. Moscow: ETS Publ.
14. Turing A. (1950) Computing machinery and intelligence. *Mind*, 59, pp. 433-460.