Эволюция искусственного интеллекта

Интервью с Александром Ждановым

доктором физико-математических наук, заведующим Отделом имитационных систем Института Системного Программирования РАН

Александр Аркадьевич, как сегодня обстоят дела с Искусственным Интеллектом?
Есть два направления развития ИИ. Одно из них заключается в моделировании тех задач, которые решает человек, или может быть — животное. Распознавание образов, вычисления, принятие решений. При этом моделируется результат решения, но не процесс. Это «прагматическое» направление, оно имеет множество программных приложений и множество успехов. Современные компьютеры вычисляют гораздо быстрее человека, и точнее. Задачи распознавания образов выполняются также на достаточно уже высоком уровне.

Второе направление — моделирование самого процесса, механизма принятия решения. Это «бионическое» направление, здесь исследуются модели живого нейрона и нервных систем. Моделирование работы мозга чрезвычайно затруднено — биология и другие естественные науки не дают полного описания работы ни мозга, ни даже одного нейрона. Существуют модели разной степени сложности, но на их основе пока невозможно приблизиться к описанию реальности, хотя кое-что уже можно воспроизвести.

Нейрон — это переключатель, или аналоговый элемент?
Очень сложный объект! И дискретный, и аналоговый. Даже простые модели нейронной сети, так называемые «искусственные нейронные сети» (ИНС), могут обучаться распознавать образы. Отличие их от природных нервных систем уже в том, что при разработке ИНС необходим кто-то, кто должен установить значения синаптических весов,

добиваясь нужного результата на выходе. Вряд ли в природе существует такой «учитель» для природного мозга. Нам же представляется, что способностями к самообучению обладает даже отдельно взятый нейрон. В этом случае у нейрона должна быть память, способность накапливать статистику и классифицировать информацию, отбирать, распознавать нужную ему информацию.

Как может быть реализована память в живом нейроне?

Я не биолог и могу говорить лишь о модели, о том, что память нейрону нужна. Как организована память in vivo, мы можем только предполагать. Может быть — пространственные характеристики синапса, или — концентрации медиаторов, или плотность каких-нибудь рецепторов на поверхности синапса, возможно — что-то в хромосомах, или все вместе.

Какие мирные отрасли в большей степени заинтересованы в разработках ИИ?

управляет системой. Поэтому, прежде всего, необходимо иметь то, чем мы управляем, какое-нибудь тело. Сама по себе, отдельно от тела, такая управляющая система — искусственный мозг, лишена смысла.

Пусть мы смоделировали мозг, например, коровы. Он очень сложен. Но какая польза в том, что мы это сделали? Какая может быть польза от искусственного мозга искусственной коровы? Ведь такой искусственный мозг должен будет иметь собственные цели, собственную волю, и собственные эмоции. В этом случае работать с таким компьютером привычным нам способом будет невозможно. Вы ему скажете: «Переведи текст», а он: «Не хочу. Я устал. Мне надо погулять по травке».

Однако задача познать, как работает мозг, это фундаментальная задача, и ее надо обязательно решать. Во-первых, просто необходимо понять, наконец, как работает мозг. Это надо, как минимум, и врачам, и компьютерщикам. Во-вторых, такого рода системам можно будет найти и практически полезную для

Если мы говорим о «бионическом» направлении, то совсем не очевидно, что искусственные системы, точно воспроизводящие мозг, кому-то очень нужны. Давайте представим себе, что мы, наконец, добились высочайших достижений, и полностью воспроизвели функции, работу живого мозга какого-то реального организма. Но в живом организме мозг является только частью системы, органом, который

людей сферу применения. Например, в таких трудных физических или временных условиях («слишком быстро» и «слишком медленно»), в которых не может работать человек. Пусть живет там, приспосабливается, накапливает знания о среде и потом передает их своим потомкам, а заодно и нам, людям, по телеметрии. В своей группе мы как раз и ставим себе целью создание такого рода систем, которые хотелось бы

называть «системами автономного искусственного интеллекта» (АИИ).

Системы АИИ мы хотим противопоставить современным «прагматическим» системам ИИ, которые есть не более чем подчиненные нам электронные рабы. Ведь такие системы ИИ не борются за свое выживание, не имеют своих собственных целей и своих эмоций. Они моделируют только третьестепенные и даже побочные функции мозга, например — умение делать арифметические вычисления, или переводить тексты. Но именно эти функции востребованы сегодня.

Согласитесь, человек был всегда заинтересован в изготовлении раба, который не имеет своей воли, а выполняет волю хозяина, в ущерб собственным интересам. Современный компьютер — разновидность раба. Мы бы не хотели, чтобы компьютер имел собственную волю, и эмоции. Поэтому, когда говорят, что хотят создать ИИ как модель живого мозга, то лукавят — не модель живого мозга хотят разработать, а электронного раба, который без устали решал бы НАШИ задачи, а не свои собственные — выжить, познать окружающую действительность, размножиться.

Человек был всегда заинтересован в изготовлении раба, который не имеет своей воли, а выполняет волю хозяина, в ущерб собственным интересам. Современный компьютер — разновидность раба

А прикладные задачи — их всегда большое множество. Везде, где есть управляемая система, мы заинтересованы в том, чтобы передать функции управления кому-нибудь более быстрому, честному и предсказуемому — компьютеру-рабу.

А зачем нужны системы АИИ сейчас?
Мы бы с удовольствием запустили такой объект с АИИ в шахту или на дно океана. Он бы добывал что-то полезное, и жизнь людей не подвергалась бы опасности.

Объекты с АИИ могли бы ползать по Марсу, приспосабливаться, собирая ценную для себя и для нас информацию. Однако обязательно надо подчеркнуть, что сейчас специалисты еще очень и очень далеки от создания такого рода систем АИИ, которые хоть как-то походили бы на искусственных животных, способных к автономному адаптивному управлению. Сегодня создаются еще только очень простые модели. На уровне бактерии, насекомого, червя... И простые организмы успешно решают задачу адаптивного управления.

А реально по Марсу сейчас ползают системы ИИ, управляемые либо с Земли по телеметрии, либо с жесткими, заранее определенными сценариями поведения, без заметных свойств адаптации, автоматического накопления знаний, изменения поведения. Они действуют по принципу «чего изволите?» и готовы безропотно закончить свои дни в одном из кратеров...

То, о чем мы говорили — об ИИ-рабе, не слишком похоже на интеллект. Это программа, сценарий. Что позволяет сказать: эта программа интеллектуальна, способна к адаптации?
Биологические системы очень сильно отличаются от машин. Организм начинается из одной клетки, постоянно развивается, изменяется, дает потомство, передает ему свой опыт и удачные генетические признаки, стареет, умирает, освобождая место потомству.

Представьте себе автомобиль, который вначале был маленьким, потом вырос, научился ездить, сдал на права и в какой-то момент времени — принес потомство. Потом он постарел, заржавел, стали отваливаться какие-то части, и потомство отвезло его на свалку, или подвергло регенерации. Делая машины, мы применяем сегодня совершенно другие принципы, несвойственные живому. Также и с современными программами. Мы обычно делаем программу, которая готова к работе в определенных условиях, она может облегчить человеку жизнь. Для того чтобы она смогла выполнять новые функции, мы должны написать новые блоки. Мы должны ее адаптировать к новым условиям работы.

А более совершенный искусственный интеллект — «автономный искусственный интеллект», он адаптируется сам, в процессе накопления информации, полученной извне.

Наше направление исследований — системы «автономного искусственного интеллекта» (АИИ). Если нам удастся отстоять и развить это направление, мы были бы счастливы.

Где могут применяться системы АИИ?
Представьте себе прибор, любое устройство, которое приспосабливается к пользователю. В настоящей жизни мы пользуемся приборами, которые адаптируем своими руками. Вы можете установить переключатель в нужное положение, или покрутить регулировки сиденья в автомобиле. Ни один из бытовых приборов не адаптируется сам к своему пользователю, как, например, домашнее животное.

Тот замысел, который воплотил конструктор, инженер, полностью определяет те пределы, в которых вы можете адаптировать прибор к себе.

В случае развития и применения адаптивных технологий прибор становится более гибким и привлекательным, более экономичным и эргономичным.

Если бы для вас исполнял сонату на пианино ваш друг, он бы старался вам понравиться, доставить удовольствие. Но если запись воспроизводит аппарат, то он не пытается настроить тембр, пространственное звучание так, чтобы вам было приятнее.

Хороший артист чувствует обратную связь, подбирает репертуар, играет для публики.

Выходит, основная цель — оптимизация.
Повышение экономичности аппаратов, освобождение от необходимости их настраивать вручную, тратить на это ресурсы и время — вот такая цель.

Многие не понимают: «Зачем это нужно, мы можем покрутить ручки, настроить, приспособить». Это результат всего развития техники 20 века, мы старались уйти от проблемы адаптации, всячески стандартизуя приборы, аппараты и технологии.

А в природе все живое адаптируется друг к другу и к неживому тоже, постоянно меняясь. Ничто живое в природе не избегает адаптации. Я думаю, разработка адаптивных систем управления — это интересное и привлекательное направление.

Какие критерии позволяют отнести систему к адаптивным? Если нужно принять решение в случае дефицита исходных данных, то человек делает эмоциональный прогноз, опираясь на «субъективное определение вероятности». Что делает машина?
В программе, обладающей АИИ, есть блоки памяти, подсистемы, которые формируют новые образы, распознают уже известные системе образы, есть база знаний, блок анализа поступающей информации. Обязателен блок принятия решений. Адаптивные функции распределены по всем этим блокам. Очень важный момент заключается в том, что присутствует аппарат эмоций, это отличает нашу концепцию «Автономного ИИ». Один из моих студентов назвал этот аппарат «Хорошометром». В каждом живом существе есть нечто, заставляющее двигаться, побуждающее к движению. Только в движении можно получить опыт и знания. Я применяю метафору батарейки, постоянно принуждающей нервную систему к действиям. Мы сейчас не обсуждаем с вами, откуда она берется, от Бога, от Природы, или появляется в процессе эволюции.

Аппарат эмоций — это некий обман, который придумала природа, чтобы заставить нас двигаться. Возбуждение отрицательного полюса аппарата эмоций вызывает негативное ощущение — ощущение боли. Положительный полюс возбужден — возникает ощущение удовольствия. Как только в систему добавлен такой аппарат, наш «организм» начинает «избегать» боли и «стремиться» к удовольствию. В норме живой организм не может получить удовольствие изнутри, из самого себя, он должен обязательно взаимодействовать со средой — и получить извне сигналы, которые будут оценены приятными эмоциями.

Вот если мысленно вытащить у живого организма, у человека, или зверюшки ту самую батарейку — аппарат эмоций, то он остановится, перестанет двигаться и со временем умрет в состоянии полного безразличия, или, может — блаженства, мы не узнаем об этом.

Поэтому природа и придумала этот обманный аппарат для живых существ.

А почему обманный?
Потому, что он заставляет начать движение и искать в окружающем мире информацию.

«Хорошометр», вроде бы, не похож на измерительный прибор. Если мы пропускаем одинаковый ток через амперметр, он показывает одинаковое значение силы тока. А «хорошометр» показывает диапазон значений, которые от одного и того же сигнала, стимула изменяются во времени. Если сегодня он показывает «5», то завтра — «4», его вообще может иногда «зашкалить».
Да, действительно, система адаптивного Автономного ИИ «живет» во времени. Живой человек все время растет, изменяется. Формируются новые образы, связи между образами и действиями. Формируются новые эмоциональные оценки для этих образов. Через год — это уже измененная система, добавились знания, образы и оценки, и решения будут приниматься несколько иные.

Важнейший критерий для системы АИИ — это ее развитие во времени.

Как в случае человека?
Два разных человека распознают одну и ту же ситуацию по-разному, и отличаются друг от друга в ее оценке. Я считаю, что большинство людей видят только то, что понимают, а что понять не могут — того буквально не замечают. Это связано с тем, что необученные нейроны просто не пропускают информацию дальше, вглубь — в область сознания.

Еще говорят, что человек видит то, что хочет увидеть.
А то, что хочешь увидеть — это ведь связано с тем, что уже является хотя бы отчасти известным, понятным. Особенно это касается деталей и мелочей. Наше поле зрения неравномерно, чтобы заметить некоторую деталь, нужно распознать, узнать ее и сконцентрировать взгляд на этой детали. А распознаем (узнаем) мы только то, что уже нам известно.

Когда я принимаю решения, я соотношу их с моей базой знаний, с тем, что я помню, а помню я закономерности. Хотя иногда закономерность, которую мы наблюдаем, является на самом деле случайностью, откуда иначе появились бы предрассудки?

Интеллект набирает базу знаний по принципу закономерности, статистической повторяемости.

У человека в памяти, в его базе знаний всегда закрепляются положительно окрашенные закономерности.
По-моему, запоминаются либо закономерности, либо прецеденты с сильными эмоциональными оценками. Но если вы повторяете последовательность действий, принимаете решение, и не получаете такого же, как и прежде результата, то данное знание, видимо, исключается из базы знаний, забывается.

Более совершенный искусственный интеллект — «автономный искусственный интеллект», он адаптируется сам, в процессе накопления информации, полученной извне

Как вы реализуете системы Автономного ИИ?
Мы, наши сотрудники, аспиранты, студенты, работаем в этом направлении более девяти лет, и у нас есть результаты, которые нас радуют и поддерживают. Это не только модельные программы, но и практические приложения — системы АИИ, которые работают на основе нашего метода автономного адаптивного управления.

Например, мы разрабатывали совместно с НПО Лавочкина программный прототип адаптивной системы для стабилизации ориентации спутника в пространстве. Стандартная система управления основана на подробном математическом описании спутника, как физического тела. Но спутник обладает множеством нелинейных осцилляторов: антенны, батареи солнечные, сам он не абсолютно жесткое тело, поэтому полностью корректную модель описать невозможно. Все коэффициенты для модели невозможно получить в точности, ведь в космосе и невесомость, и резкая смена температур, вакуум, воздействие магнитных сил.

Мы разработали программу, но спутник, для которого она была необходима, так и не полетел. Он должен был нести рентгеновский телескоп, и как раз для него была необходима точная ориентация на объект наблюдения (какую-нибудь звезду, например).

Другой пример — мы сделали на заказ фирмы ATS прототип адаптивной системы управления активной подвеской автомобиля.

Это как у Ситроена?
«Активные подвески» есть у Мерседеса, у Ситроена, заявлены и другими фирмами. Например, Ситроен предлагает две-три опции, которые переключает водитель. Это, конечно, «адаптация», но слабая. Наша активная подвеска «собирает» информацию о состоянии подвески и о загрузке автомобиля, формируется база знаний, и программа передает команду исполнительному устройству, которое действует на активный амортизатор, и удар от камешка или ямки

компенсируется. Автомобиль (его адаптивная система управления подвеской) после начала движения как бы «учится», и в продолжение движения «применяет» накопленный опыт.

То есть водителю в начале движения необходимо наехать на пару бордюров, и попасть в колодец?
В наших условиях все особенности дорог не заставят себя долго ждать.

Несколько лет назад была нами разработана и прототипная система поддержки принятия решений при управлении социальными объектами для исследовательского подразделения Администрации Президента РФ. В качестве демонстрации возможностей нашей системы адаптивного управления мы рассмотрели социальный объект, который можно было описать рядом факторов, история их изменения была известна для ряда управляющих воздействий. Применение такой адаптивной системы на моделях позволяло смягчить проявление нежелательных эффектов. Например, вовремя начать реагировать при появлении признаков опасной ситуации, или прекратить воздействие на систему тогда, когда это стало не нужно, тогда, когда система регулируется сама по себе. Любое административное воздействие — это расходы, и если применяются ненужные, лишние воздействия к самоуправляющемуся, саморегулирующемуся социальному объекту, мы тратим средства зря.

Были ли в процессе исследований такие ситуации, когда возникало ощущение чуда, когда вы или ваши коллеги были удивлены тем, как работает программа, или программно-аппаратный комплекс, или, как говорится, «все под контролем»?
Было. В программе со спутником. Я сначала отмечу, что начало процесса обучения Автономного ИИ по времени должно совпадать с началом действия того объекта, которым он управляет. Нет такого, что сначала — учись, а потом — живи. Обучение и управление в принципе должно быть «в одном флаконе», и мы в своих системах это реализуем.

Это была бы идеальная образовательная система.
Да, хотя в реальной жизни эти процессы часто разделяются, за счет смещения акцентов — то в сторону задачи обучения, то в сторону задачи управления. Например, в нашей жизни акценты расставлены так, что в детстве ты больше обучаешься и меньше управляешь, а потом — больше управляешь и меньше обучаешься.

В процессе «обучения» управляющая система совершает пробные воздействия. И объект откликается по-разному. Однажды при работе с системой управления угловым движением спутника мы выяснили, что при грубом и резком воздействии на модель спутника, она вела себя иначе, чем при мягком и бережном. Оказывается, свойства объекта зависят от того, как вы с ним обращаетесь. И в жизни так. Если с человеком обращаться помягче — результаты лучше, чем если вы его трясете, как грушу.

Еще пример. Была программа, имитирующая обучение езде на велосипеде. Вначале «мальчик» в программе падал, и не мог удержаться на полосе движения. Потом он нашел способ ехать не падая, но при этом совершал сильные колебательные движения рулем, достаточно долго, и только потом система управления («мальчик») нашла способ ехать прямо при очень экономных движениях руля. При этом мне вспомнилось, как-то раз друг повел мой автомобиль, у него большой водительский опыт, но моя машина была новой для него, она была больше. И он вначале ехал достаточно ровно, но совершал при этом те же самые колебательные движения рулем, которые прекратились только лишь через некоторое время. Это было удивительно — совпадение естественного и моделированного поведения.

В случае системы с социальным объектом, мы «обучали» ее на архивных данных, на определенном промежутке времени. Потом получили график рекомендованного управляющего воздействия для всего времени архива. Я показываю график заказчику, и говорю:

— Вот здесь, в такой-то момент, наша программа посоветовала бы действовать более активно.

— Да, тут с решением мы немного опоздали, можно было бы избежать нежелательного ухудшения процесса.

— А вот здесь, позже, вообще нужно было бы отказаться от воздействий на объект.

— Я же им говорил, говорил, что ничего не надо делать, все само образуется! — вскричал мой собеседник.

Итак, есть два направления: первое — изучение мозга, и последующие попытки воспроизведения его работы; второе — построение ИИ под задачу, с целью ее решения. Какое направление сейчас преобладает?
Конечно же, количественно преобладает прагматическое направление — промышленное производство электронных рабов — которые есть не просят, и 24 часа в сутки считают, обрабатывают, передают почту, принимают стандартные решения.

Скажите, возможно ли эмоционально окрашенное отношение к машине?
Если это будет система Автономного ИИ, обладающая памятью событий своей жизни, эмоциями, то вы ее

будете беречь, хотя бы потому, что от значений эмоциональных оценок образов зависит, какие решения будут приниматься системой (так при воспитании детей надо особенно заботиться о том, чтобы у них сформировались правильные критерии — правильные оценки объектов, событий, это будет влиять на их предпочтения потом, когда они будут принимать решения).

Но уже и сейчас иногда встречается антропоморфное отношение к компьютеру: «Он устал, давайте выключим его, чтобы не грелся».

Министр Иванов как-то раз сказал, что расширение НАТО в Европе из угрозы превратилось в опасность. Различие между угрозой и опасностью и суть понятий ощутимы. Есть ли какая-нибудь опасность в том, что большинство людей, которые пользуются PC, на самом деле не понимают, как все это устроено, как это все функционирует?

Если ты не понимаешь, как все устроено, то ты не можешь понять и опасностей, которые могут возникать. Есть ли опасность в том, что огромное число людей — пользователей PC адаптируется именно к выдуманному Биллом Гейтсом интерфейсу? Я не знаю, не могу ответить.

Некоторые опасности можно указать уже сейчас: человек приспособлен к взаимодействию с предметным миром вещей и миром людей. Вот я вижу кружку, беру ее и пью кофе. Вот я вижу опасность — я стараюсь убежать от нее. Вот — женщина, я могу поговорить с ней. Вот — студенты, я буду читать им лекцию. Каждый объект в реальном мире порождает мой способ взаимодействия с ним, в соответствии с моим опытом и желаниями, с поставленными целями. В случае компьютера, я вижу кружку кофе — но на экране дисплея, я хочу пить, но не могу взять эту виртуальную кружку. Максимум, что я могу сделать — это нажать на кнопку.

Телевидение, компьютеры разрывают обратную связь с миром. Вас либо совсем лишили возможности действовать, реагировать на видимую картину, либо оставили только одну возможность действия — нажимать на кнопки. Вы участвуете в игре, видите битву, вы переживаете, у вас вырабатывается адреналин, вы должны вскочить и физически участвовать в действии, а вместо этого вы остаетесь сидеть в кресле. Максимум, что вы можете сделать — это нажать на кнопку. Я не врач, но могу предположить, что от этого тело как-то пострадает. Второй вариант — вы адаптируетесь, и информация на экране перестанет вызывать у вас эмоции, и не будет приводить ни к каким действиям. Если время жизни в виртуальной реальности растет, то вы и в обычной деятельности приучитесь не испытывать эмоций, к чему это приведет? Вот такая опасность. Здесь главное — правильное соотношение времени, проведенного в виртуальной и в естественной реальности, и качество переживаний.

Если разорвана обратная связь системы Автономного ИИ с управляемым ею объектом, что происходит на модели?

А ничего хорошего не происходит. Система перестает развиваться, накапливать базу знаний и опыт, и, в конце концов, становится неспособной к управлению, к той «жизни», для которой она предназначена.

Если мысленно вытащить у живого организма аппарат эмоций, то он остановится, перестанет двигаться и со временем умрет в состоянии полного безразличия, или, может — блаженства, мы не узнаем об этом

Какие вы видите пути развития технологии виртуальной реальности (ВР)?

Надо обязательно искать полезные способы ее применения. Например, тренажеры для летчиков, водителей, операторов различных машин (хотя и в этих приложениях обнаруживаются свои опасности). В образовании. В системах, моделирующих какие-то уникальные ситуации, для принятия более правильных решений. Способы использования ВР в лечебных целях в медицине, в протезировании.

Развивать ВР просто для развлечения, по-моему, очень опасно. Если заменить всю афферентную, входящую сенсорную информацию на искусственные сигналы, да еще связать их с эфферентными, выходными нервными сигналами человека, то это очень опасное вторжение в нервную систему человека, в его психику, это путь к распаду человека. Плохо, что это естественное и мощное желание, которое уже трудно сдержать. Это путь к подмене реального мира искусственным, которым можно легко манипулировать, путь, начатый еще театром, затем кинематографом, телевидением, потом продолженный компьютерами...

Вы правильно связали здесь системы ИИ, АИИ и ВР. В каком-то смысле это одного поля ягоды — это системы искусственной жизни. Вероятность их использования во вред очень велика, впрочем, как и любых результатов технического прогресса. Однако видна и большая польза, которую могут принести такие системы. Еще Парацельс говорил: «Все есть яд и все есть лекарство, важна лишь мера!» Давайте будем бороться за меру и пользу. Ψ