



*Воинов Александр Владимирович  
Гаврилова Татьяна Альбертовна  
Грановская Рада Михайловна*

ОБСУЖДАЕМ ГЛАВЫ ИЗ КНИГИ  
**МАРВИНА МИНСКОГО и ГАРРИ ГАРРИСОНА**  
**"ВЫБОР ПО ТЬЮРИНГУ"**  
(ПЕРЕВОД В.А. МОИСЕЕВА)

*Развитие информационных технологий происходит так стремительно, что в школьный курс информатики не успевают попадать наиболее современные и перспективные идеи из области компьютерных наук. Поэтому мы сочли важным ввести рубрику, в которой эти идеи излагаются доступным языком в виде научно-популярных статей или даже научной фантастики. Для первого опыта мы выбрали публикацию по искусственному интеллекту. В основу статьи положены материалы неопубликованных глав научно-фантастического романа "Выбор по Тьюрингу", принадлежащего перу двух известных людей: специалиста в области искусственного интеллекта Марвина Минского и писателя-фантаста Гарри Гаррисона. Книга написана на основе фундаментального труда Марвина Минского "Общество разума" ("Society of Mind"), который, к сожалению, не переведен на русский язык. В этой книге Марвин Минский объясняет, как "неумеющие мыслить" части могут образовать мыслящее целое.*

*Цитируемые ниже главы были исключены из основного текста романа как замедляющие действие, но зато они чрезвычайно интересны как научно-популярная статья. Перевод этих глав блестяще сделал писатель-фантаст Владимир Анатольевич Моисеев.*

*В романе фигурируют как математики и программисты (специалисты по искусственному интеллекту), так и психологи, поэтому мы попросили прокомментировать цитируемые главы отечественных специалистов по искусственному интеллекту и психологии. Психологические комментарии сделаны доктором психологических наук, профессором Санкт-Петербургского университета Р.М. Грановской, послесловие написано доктором технических наук, профессором кафедры "Компьютерные интеллектуальные технологии в проектировании" Санкт-Петербургского технического университета, руководителем городского семинара по искусственному интеллекту Т.А. Гавриловой, а вступительная статья подготовлена сотрудником Института прикладной астрономии РАН, кандидатом технических наук А.В. Воиновым.*

*Эти статьи показывают, насколько в современном мире сблизились проблемы технические и проблемы психологические. Для сравнения обратите внимание на продолжающуюся публикацию факультативного курса "Система автоматизации процесса решения задач" авторов М.Дмитриевой и М.Павловой, который показывает, какой содержательный математический аппарат стоит за исследованием проблем искусственного интеллекта.*

*Итак, предоставляем слово А.В. Воинову.*

**Платон и Сократ грозили ненавистникам разума всякими бедами. Но дана ли им была власть оберечь от бед того, кто разум возлюбит?**

*Л.Шестов*

**Вопрос «может ли компьютер думать?» ничуть не более интересен, чем вопрос «может ли подводная лодка плавать?»**

*Эдсгер В. Дейкстра<sup>1</sup>*

**Гниет здесь гордая латынь. Аминь.**

*Эпитафия Высокоученому (автор неизвестен)*

## **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ. ИСКУССТВЕННАЯ ДУША?**

### **ВВЕДЕНИЕ**

Есть класс задач, которые принято называть «интеллектуальными». Для их решения от человека требуется то, что принято называть «интеллектом». «Искусственный интеллект» – это наука о том, как сформулировать «интеллектуальную» задачу так, чтобы ее решение можно было автоматизировать, то есть составить для компьютера, ничем в сущности не отличающегося от утюга или пылесоса, набор инструкций («программу»), с необходимостью приводящий к требуемому решению.

Начала искусственного интеллекта, как и многого другого, заложили греки. Гением Сократа, Платона и Аристотеля была создана логика – наука о некотором формализуемом подмножестве человеческого мышления. Ни у одного народа, современного этим великим грекам, мы не найдем такого интереса к чисто рассудочной стороне мышления, к его ясности и отточенности. Значение создания формальной логики трудно переоценить. Впервые человек получил возможность «оторваться» от своего способа думать (а ведь он мог считать, что это и есть он сам!), посмотреть на него со стороны и отнестись к нему как к объекту исследования. В принципе, в этот момент и был создан искусственный интеллект. В самом деле, как только на бумаге были выписаны формальные правила преобразования «истинных» высказываний, так они сразу же начали «жить своей жизнью». Правила на то и правила, что испол-

няются с железной необходимостью. Выписывает ли их живая рука Аристотеля, или кристалл кремния «производит» их внутри себя под управлением «интеллектуальной программы» – это уже не имеет значения. Правила вывода в обоих случаях одни и те же.

Итак, есть логика. Почти сразу же возникает ощущение, что оставшаяся, «не дотянувшая» до формализации часть мышления даже и недостойна нашего внимания. В науке такая постановка вопроса естественна и осмысленна. В самом деле, очень часто исчерпывающее решение какой-либо частной, упрощенной задачи помогает решить (или хотя бы лучше понять) ту, что, собственно, и была изначально поставлена. Кроме того, если честно и непредвзято поставить перед собой вопрос, зачем нам нужно мышление, то, может быть, и хватит того, что вписывается в рамки логики?

Пропустим блаженного Августина (и его влияние на умы Западной Европы), Фому Аквинского, схоластиков, Паскаля и пр. и перейдем сразу к Декарту. Провозглашенный Декартом принцип рационализма задавал направление развития всей европейской цивилизации, ее науки и технологии. Декартов рационализм, в интересующем нас отношении, сводится, во-первых, к рациональному, рассудочному постижению опытно достигаемого мира и основанному на этом

<sup>1</sup> Известный специалист в области программирования 60-70 годов.

постижении освоению (иначе зачем и постигать?) и, во-вторых, постижению самого человека как мыслящей машины. Второе положение очень важно, хотя мы, наверное, уже так к нему привыкли, что не в

**Декартов рационализм, в интересующем нас отношении, сводится к постижению самого человека как мыслящей машины.**

состоянии оценить его значение для тогдашней образованной Европы. Декарт поставил человека в один ряд с объектами внешнего мира и явлениями природы. Для исследования и тех и других предлагался один универсальный метод – рациональной, логической интерпретации данных чувственного опыта. Казалось бы, что здесь особенного? Принадлежность физиологической оболочки человека чувственной природе ни у кого не вызывала сомнений. Однако для Европы того времени мир души, к которому относили и способность мыслить, был еще окружен покровом таинственности.

Рационализм Декарта в «эволюционном» отношении оказался наиболее выигрышным. Сосредоточив свои силы на освоении и покорении природы посредством «научно-технического прогресса», европейские народы далеко обошли всех остальных. Победителей не судят. Хотя во времена Декарта его точка зрения не была единственной, фантастический прогресс «разума» в последующие столетия заставил за-

**Есть строгий математический факт, – теорема Геделя, – выражающий ограниченность человеческой способности составить осмысленное высказывание.**

молчать его «ненавистников». Вспомним оптимизм Жюль Верна и ранних Стругацких. Казалось, что тот «интеллект», посредством которого человек открывает уравнения механики и придумывает лампочку

Эдисона, и есть самое развитое в нем и высшее.

Таким образом, на долгие годы тот способ мышления, который не укладывается в рамки логики, с помощью которого невозможно составить формально доказательное (и, следовательно, «объективное») суждение, отошел в область «беллетристики». Афоризм – остроумен, но бездоказателен.

## **АЛГОРИТМИЗАЦИЯ И МАТЕМАТИКА.**

### **ГИЛЬБЕРТ И ГЕДЕЛЬ**

Царицей европейского рационализма безусловно является математика.

Именно в математике было в максимальной степени достигнуто умение «исключать неоднозначность из определений».

«Царственное» положение математики среди других наук в течение многих веков породило уверенность, что именно на пути полной математизации знания возможна и осмыслена также и его алгоритмизация. В самом деле, составить алгоритм решения квадратного уравнения легче, чем правило диагностики легочного заболевания, не потому, что первая задача проще, а потому, что она в большей степени формализована и переведена на язык математики.

Сильно укрепили эту уверенность успехи естественных наук, которые опираются на математику в интерпретации знания и его приведении к такой форме, которая допускает объективную, не зависящую от частных, немоделируемых особенностей исследователя проверку, а также переносимость без искажений от одного индивида к другому.

Программа полной математизации всех наук, выдвинутая Гильбертом, провалилась. Оказалось (в частности), что есть строгий математический факт, – теорема Геделя, – выражающий ограниченность человеческой способности составить осмыс-

ленное высказывание.

Тем не менее, именно в рамках той серии работ, которую вдохновила программа Гильберта, была разработана теория рекурсивных функций, тщательно исследованы вопросы формальной «вычислимости» функций и пр., без чего невозможно было бы подойти к составлению сложных алгоритмов для реальных компьютеров.

### **АВТОМАТИЗАЦИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ЗАДАЧ: НОВЕЙШАЯ ИСТОРИЯ**

Оставив в стороне докомпьютерные способы механизации вычислений и «рассуждений» (простой дорожный светофор дает пример схемы автоматических рассуждений), перейдем к компьютеру в его современном виде и программированию для него.

Компьютер – это «вычислитель». В эпоху мультимедиа и компьютерных игр трудно даже представить себе, что изначальное назначение компьютера – складывать и умножать числа.

Первым применением компьютера действительно были простые арифметические расчеты. Это была первая «интеллектуальная» операция, автоматизированная полностью и исчерпывающим образом. Кажется банальным упоминать об этом, но во многих учреждениях, особенно там, где производились огромные объемы вычислений (астрономия, навигация), был специальный штат сотрудников-«вычислителей», которые целыми днями складывали и умножали числа (на арифмометре и по таблице логарифмов). Эти сотрудники часто имели высшее образование по профилю учреждения, но реально сто процентов их рабочего времени было занято чисто рутинными операциями.

Следующей интеллектуальной операцией, также полностью автоматизирован-

ной, стали бухгалтер и хранение нечисловой информации (50-60 годы).

Третьей значительной сферой интеллектуальных операций, в которые компьютер вошел всерьез и надолго, стали аналитические выкладки. Первые программы, которые могли взять в символьной форме производную от алгебраического выражения, появились еще в 50 годы.

О шахматах и говорить нечего. Это была четвертая по счету сфера, занятие которой требует интеллекта и которая так же, начиная с 60 годов и по настоящее вре-

### **Классический искусственный интеллект**

- Теория автоматов
- Логическое программирование и автоматическое доказательство теорем
- Инженерия знаний
- Представление знаний (Минский, формализм фреймов)
- Экспертные системы
- Обучающие системы
- Автоматическое самообучение (machine learning)
- Нейронные сети
- Имитация коллективного решения интеллектуальных задач: многоагентные системы и распределенный искусственный интеллект
- Виртуальная реальность

мя, полностью и исчерпывающим образом автоматизирована.

То, что называется искусственным интеллектом в узком смысле, – разработка экспертных систем и систем поддержки принятия решений в слабоструктурированных областях знания, также начиналось в 50 годы. Как вспоминает Дж. Минкер, известный американский специалист в области логического программирования, уже тогда по заказу министерства обороны США велись разработки в сфере, которую условно можно назвать «интеллектуальные базы данных». Эти разработки с полным правом можно отнести к сфере искусственного интеллекта.

## РЕЗЮМЕ

Один из столпов современного российского искусственного интеллекта Д.А. Поспелов 30 лет назад рассматривал идею так называемой «психоники» – научного направления, которое, аналогично бionике, должно было бы «перенимать» у реального, естественного интеллекта те приемы, которыми он пользуется при решении задач и воплощать эти приемы в интеллектуальных программах.

Практика показывает (и сам Дмитрий Александрович вряд ли будет против этого возражать): способы решения задач человеком и компьютером (точнее: с одной стороны – способы, наиболее удобные и эффективные для человека, а с другой – для компьютерной программы) принципиально различны. Можно, конечно, запрограммировать точную имитацию того, как человек решает какую-то конкретную задачу, но это будет громоздко и неэффективно.

Даже в одном из актуальных в наши дни направлений, возникшем на стыке психологии и инженерии знаний, в рамках которого ставится задача о моделировании образного мышления человека и его интуиции, среди десятка самых различных подходов к решению этой задачи нет такого, который бы точно копировал какие-то свойства человеческой психики.

## СЛЕДУЮЩИЙ ШАГ: АНТРОПОМОРФНЫЙ ИНТЕРФЕЙС

Совсем недавно несколько государственных ведомств Японии собрали специальную экспертную комиссию с целью выявить те направления информационных технологий, в которых Япония может наиболее быстро и прочно выйти на лидирующие позиции. То есть именно в эти направления японское государство предполагает вкладывать свои огромные финансовые средства.

Наиболее интересным из этих направ-

лений (комиссия их выделила порядка десяти) является проект создания человекоподобного (антропоморфного) интерфейса.

Интерфейс. Слово, понятное каждому, кто хоть немного поработал с компью-

Ставится задача создания нового, «человекоподобного» компьютера, внутренне устроенного так, чтобы все задачи решать интеллектуально. Речь идет о том, чтобы вместо привычных уже многим из нас окон, кнопок и меню, мы видели бы человеческое лицо, слышали бы узнаваемую, интонационно окрашенную речь и спокойно, на естественном языке, общались бы, что нам от компьютера надо.

тером как профессиональный пользователь. Компьютер многолик. Его содержимое может быть (и часто бывает) одно и то же, но для разных категорий пользователей оно может выглядеть совершенно по-разному. Ниже мы увидим, как одна и та же интеллектуальная программа выглядела для гостя со стороны как развесистое дерево, а для ее автора продолжала оставаться набором файлов, драйверов и утилит командной строки.

В рассматриваемом японском проекте речь идет именно об интерфейсе. То есть не ставится задача создания нового, «человекоподобного» компьютера, внутренне устроенного так, чтобы все задачи решать интеллектуально. Речь идет о том, чтобы вместо привычных уже многим из нас окон, кнопок и меню, мы видели бы на экране (или в шлеме «виртуальной реальности») человеческое лицо, слышали бы узнаваемую, интонационно окрашенную речь и спокойно, на естественном языке, общались бы, что нам от компьютера надо.

## МАШИННАЯ ЦИВИЛИЗАЦИЯ... БЕЗ ЧЕЛОВЕКА?

Не только сам «плотский» интеллект, но и весь основанный на нем способ адаптации к внешнему миру посредством освоения и покорения его, как выясняется, может быть автоматизирован.

В самом деле, современная техноло-

гическая цивилизация основана на покорении детерминированно моделируемой части природы. Покоряя природу, человек ставит между собой и природой – технику. И самого себя он (в соответствии с упомянутым принципом Декарта) пытается свести к одному из фрагментов такой детерминированно покоряемой природы. Есть даже такое направление исследований – на стыке социологии и искусственного интеллекта, – которое результаты, полученные при изучении коллективных интеллектуальных систем, пытается применить к изучению способов управления и оптимизации человеческих коллективов.

Но мир техники, замкнув обратную связь на себя и внутри себя, по-видимому, вполне может самостоятельно «справиться» с природой. С той ее частью, которая доступна постепенному, поэтапному «отражению» в рациональных понятиях и вытекающих из этого отражения практических рецептов по достижению относительно природы значимых целей. А выходит за рамки этой части природы только человеческий дух. Но дух не существует для кибернетического Робина (см. ниже) и его творцов.

**Важно помнить, что тот механический мир, которым человек окружил себя, человеку органически чужд. И по плоти, и по душевному устройству (по способности думать), и по духу.**

Можно поставить эксперимент. Запустить такого Робина в джунгли и посмотреть, что потом от бедных джунглей останется... Если серьезно, можно смело утверждать, что программа, способная к самообучению и имеющая доступ к адекватной информации о внешней среде – в реальном времени – вполне справится с задачей выживания в джунглях.

### ПОМЕЧТАЕМ?

Упомянутый выше известный специалист в области искусственного интеллекта М. Минский недавно выложил на ftp-сервер Массачусетского института техно-

логии (где он работает) две неопубликованные главы фантастического романа, написанного им самим в соавторстве с профессиональным писателем-фантастом Г. Гаррисоном.

Представляется целесообразным отнестись к этому повествованию как к некоторому «мысленному эксперименту», моделирующему наше ближайшее будущее и проанализировать, во-первых, наши перспективы, а во-вторых, точность, «надежность» и «валидность»<sup>2</sup> прибора: того, как видит эту ситуацию автор – американский профессор, репрезентативно отражающий мировоззрение тех людей, усилиями которых предлагаемая «сказка» уже практически становится «былью».

Прежде всего, отметим, что это уже не фантастика. См. выше про антропоморфный интерфейс. Голосовой ввод информации в компьютер, распознавание визуальных образов, синтез речевых сигналов и т.д. уже давно существуют.

Что касается распознавания («понимания») естественного языка, то сам Минский внес значительный вклад в разработку формализмов, которые приближают решение этой задачи. Здесь вопрос не в качестве, а в количестве, и нет сомнения, что японские инженеры с присущей им генетической тщательностью доведут дело до конца.

Автор, не знакомый с понятием духа и его личностной (ипостасной) уникальностью, фактически смешивает эту уникальность со способностью удерживать изменение своих взглядов в каких-то заданных извне рамках.

Автор прибегает к теории Фрейда, так как, по-видимому, эта теория наиболее известна и доступна на уровне массового сознания. Мы не собираемся «ругать» Фрейда, напротив, он гениально увидел в человеке многое из того, что необходимо видеть врачу-психиатру или психологу, что-

<sup>2</sup> Термины экспериментальной психологии.

бы реально этому человеку помочь. Но всякий метод работает тем лучше, чем лучше известны его границы. По-видимому, многие психологи согласятся с тем, что в рассматриваемой ситуации теория Фрейда наименее адекватна. В самом деле, сильная сторона теории Фрейда заключается в том, что она отражает те особенности человеческой психики, которые испытывают максимальное влияние со стороны его «hardware» – (грубо говоря) физиологии. И наоборот, она наиболее слаба в той сфере, которую игнорирует – в сфере духовной.

Тот способ имитации человеческих страстей (инстинктов и пр.), который описан в цитируемых главах и который, по мнению автора, должен помочь роботу, во-первых, выглядит слишком поверхностно даже для фантастики, а во-вторых, зачем роботу усугублять проблемы его собственного природного устройства имитацией чужой природы?

Далее, герои повествования последовательно переносят в компьютерную программу то, что они считают присущим человеческой природе, с тем, чтобы получить «существо», которое вело бы себя в реальном мире не менее осмысленно, чем человек.

Но можно ли утверждать на сто процентов, что нравственное начало присуще природе человеческой? Многие миллионы людей на Земле уверены как раз в обратном: в том, что это начало привнесено в человека извне, и мы не можем, в частности, по причине полного отсутствия объективного знания в данной области, сбрасывать со счетов эту многовековую уверенность. В самом деле, многие люди, поступая нравственно, могут, как Робин, объяснить (рационализировать по Фрейду) свои

поступки соображениями целесообразности, выгоды и т.д. Но с другой стороны, очень часто нравственное поведение человека идет совершенно вразрез с его природными потребностями и соображениями даже отдаленной перспективы его адаптации.

Человеческая природа очень несовершенна. Человек склонен совершать и сознательно злые и просто неразумные поступки. Тем не менее, упомянутое «нравственное начало», безотносительно к его источнику, удерживает его, и доброго и злого, в некоторых «разумных пределах». Что послужит Робину аналогом этого начала, если человек, не будучи его хозяином в себе самом, не сможет его смоделировать в интеллектуальной программе? Мысленный эксперимент Минского оставляет этот воп-

**Фантастика и реальность. Во многих произведениях фантастов фигурировали кибернетические (домашние) животные, и обычно это выглядело весьма безобидно. Но вот появился «тамагочи», и мы вздрогнули.**

рос открытым. Очень скоро (в ближайšie десять-пятнадцать лет), тем не менее, реальный эксперимент

покажет, какие из сторон феномена человека доступны для компьютерного моделирования.

Данная статья не должна рассматриваться как призыв что-то запретить, осудить или заклеить. Напротив, развитие искусственного интеллекта и прогресс основанных на нем технических устройств все более помогает нам понять, что же все-таки является существенно человеческим?

Кроме того, и это самое важное, рассмотренный мысленный эксперимент действительно рисует наше ближайшее будущее. Мы вступаем в XXI век. Мы уже очень мало можем повлиять на то, каким он будет. Но нам в нем жить. И надо быть готовым к тому, чем он встретит нас и наших детей.

*А теперь перейдем к обсуждению неопубликованных глав романа "Выбор по Тьюрингу". Вот что пишет Марвин Минский об истории этих глав:*

Гарри Гаррисон и я давние друзья. Как-то он сказал мне, что ему очень понравились идеи, содержащиеся в моей книге "Общество разума". Но чтобы они смогли найти распространение среди большей аудитории, мне стоит написать более популярную версию в форме романа. Я ответил, что не обладаю необходимым для такой работы талантом. Гарри предложил свою помощь.

Мы договорились, что главным персонажем книги станет математик суперхакер из будущего, который сумел создать первый искусственный интеллект и придал ему человекоподобное мышление. Гарри взялся обеспечить наш замысел действием, от меня требовались технические подробности, причем в большом количестве. Я восхищался работой Гарри по крайней мере 20 лет. Литературный агент Гарри продал план в "Уорнер Букс", и мы потихоньку начали работу над книгой. Через пару лет появился замысел сюжета, в котором герой получает серьезную мозговую травму. Это позволило нам объяснить вычислительную часть теории на примере восстановления мозга Брайена, одновременно рассматривая психологические коллизии, возникающие при возвращении его детских воспоминаний.

Соавтор не имеет права распоряжаться своей работой самостоятельно. Эти две главы - часть текста, написанного мной, но так и не вошедшего в окончательный вариант книги. И Гарри, и Брайен Томсен, редактор из Уорнер Букс, посчитали, что они чересчур замедляют действие. И все-таки я планирую поместить оставшуюся часть материала в другую книгу, посвященную тем же проблемам - единственное название, которое кажется правильным, было бы "Робопсихология" - термин, введенный в свое время Азимовым. Мы надежались показать Айзеку проект главы, касающейся сохранения робота (или человека) от надвигающегося безумия, но в прошлом году он был настолько болен, что мы не хотели тревожить его."

24 сентября 1992 года.

*Познакомимся, наконец, с основными идеями Марвина Минского, сформулированными в романе и прокомментированными Р.М. Грановской.*

19 июня 2024 года

Когда Брайен и Бен достигли лаборатории, компьютер был включен, но робот был неподвижен.

- Робин, включайся.

Робот-дерево повернулся к ним и произнес:

*Р.Г. Робот-дерево - это интересная конструкция тем, что, с одной стороны, много степеней свободы для рук, (руки ветки с тысячами крошечных пальчиков), а с другой, нет базовой подвижности. Очень похоже на человека, укоренившегося возле компьютера.*

- Доброе утро, Брайен. И незнакомец.

- Это - Бен. Ты встречал его раньше.

- Доброе утро, Бен. Предполагаю, что мы, наверное, встретились впервые совсем недавно, потому что я не нахожу никакой информации о вашем появлении в записях своей постоянной памяти, в которой хранятся события, случившиеся более месяца назад.

*Р.Г. Действительно, трудно понять, как человек догадывается о том, что для него уже когда-то было, а что, действительно, - впервые. Здесь описана одна из гипотетических возможностей: робот сканирует свою долговременную память. Однако, это выглядит неправдоподобно, поскольку перебор постоянной памяти - это надежная, но медленная процедура. Обычно, у человека, это делается иначе и связано с эмоциональным ощущением знакомости. Конечно, последнее не исключает ошибок, но работает - быстрее.*



- Ты прав, Робин.

Бен повернулся к Брайену.

- Проклятье! - пожаловался он. - Хотел кое-что записать, но у меня закончилась бумага.

- Даже предположить не мог, что кто-то до сих пор использует этот материал, - проворчал Брайен, - Робин, пожалуйста, принесите Бену несколько листов бумаги.

- Сколько листов я должен принести? - спросил Робин.

- Ну, несколько.

- Я не понимаю, что вы подразумеваете под словом "несколько"?

- Несколько - указывает на небольшое число, такое, как три или четыре.

*Р.Г. Здесь затронут один из самых трудных вопросов. На самом деле для определения понятия "НЕСКОЛЬКО" требуется использовать конвенциональные критерии. Обычно, в рамках европейской культуры, при этом подразумевается число в окрестности 10. Таких условностей у человека великое множество, и он не замечает, что они некорректны. Например: "Я ЗАЙДУ ПОЗДНЕЕ" или "МЫ ЗАЙМЕМСЯ ЭТИМ В ДРУГОЙ РАЗ". В обиходе подобные сообщения вполне приемлемы, а вот их уточнение - слишком дорого.*

- Я понял. Мне следует принести три или четыре листа?

- Нет, Робин, Ты не понял. Выражение "три или четыре" является идиомой. Ты знаешь, что такое идиома?

- Да, сэр. Идиома - это языковое выражение, чье значение исключительно и не может быть получено из соединенных значений элементов, или стиля артистического выражения, или структурной характеристики формы некоторого ...

*Р.Г. Здесь считается, что правил построения идиомы не существует. Это не совсем так. Формирование идиомы связано с правилами построения МЕТАФОРЫ как обобщенного портрета при данном ракурсе рассмотрения проблемы (см. лит. 1, 2, 3, 4, 5).*

- Подожди, Робин! - вмешался Брайен. - "Три или четыре" всего лишь способ указать на число, близкое к трем или четырем.

- Я понимаю. Я принесу 99 листов.

- Но почему ты выбрал именно 99?

- Потому что "близко к" понятие приблизительно, - ответил робот. - И 4 близко к 5. И 5 - почти 6. И 6 - почти 7. Со-

гласно моему словарю, я использую эти слова правильно. И 8 - приблизительно ...

- Подожди, Робин! - Брайен повернулся к терминалу и проверил программу. - Ага. Беда Робина в заложенной в него концепции аппроксимации. Декларативное определение кажется в порядке, но процедурное определение дефектно.

- Что это значит?

- Одно дело знать, как термин определен - и совсем другое - уметь применять свое знание на практике. В частности, нельзя использовать метод аппроксимации несколько раз, как это сделал Робин. Погрешности накапливаются слишком быстро. Вот и Робин, без малого, застрял в бесконечной петле.

*Р.Г. Здесь важно иметь программу трансляции команды из ЛЕВОГО полушария в ПРАВОЕ. При подобной трансляции неизбежно теряется однозначность определения, но возникает однозначность последовательности исполнения.*

- Но почему этот процесс вообще остановился, когда достиг 99?

- Хмм... Теперь, когда вы спросили об этом, я действительно не знаю. Давайте посмотрим выполнение предыдущих команд.

Брайен вернулся к терминалу.

- О, взгляните на это. Сначала он пришел к заключению, что "несколько" может быть определено как "не много". Затем обнаружил пример, где сотня предметов называется многими. Вот он и использовал одну сотню как проверку на принадлежность к понятию "не много"!

- Это будет легко исправить?

- Должно быть. Я только поручу Б-мозгу удерживать А-мозг от повторения подряд одного и того же действия.

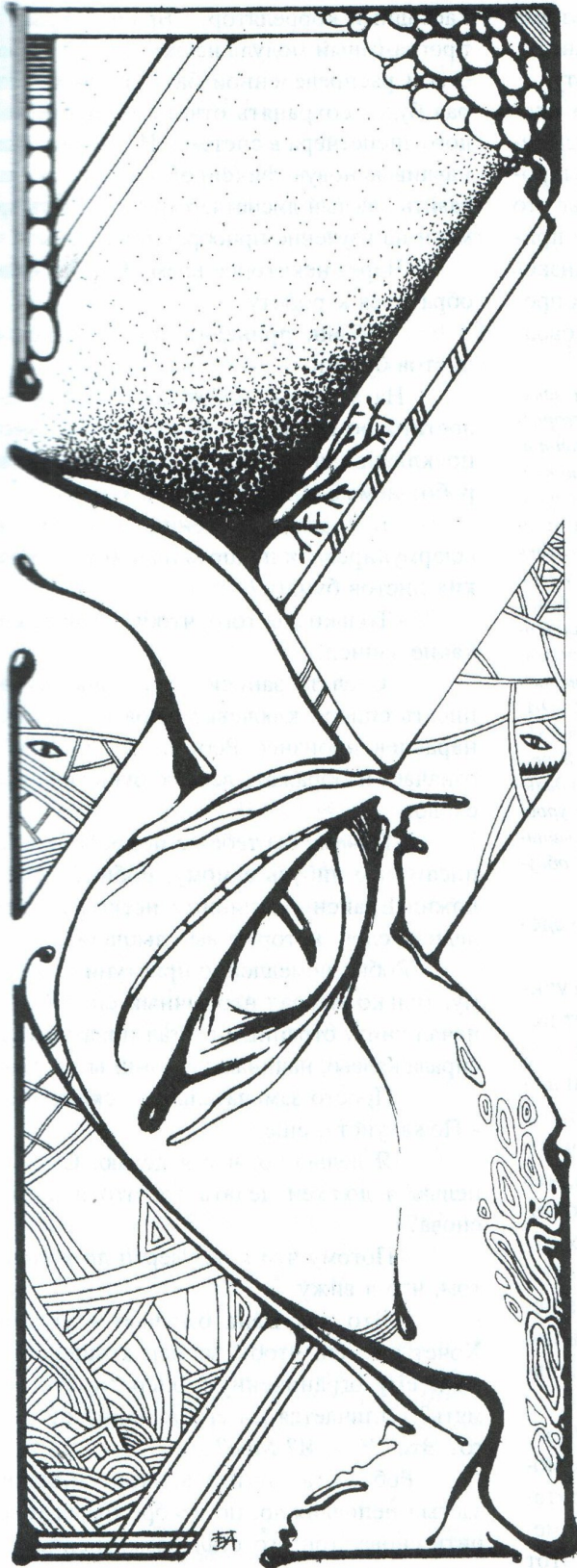
Через минуту Брайен поднял глаза.

- Он должен это делать. А теперь, Робин, принесите немного бумаги для Бена.

Робин начал пересекать комнату, но внезапно повернул налево, совершая все более и более сложные движения.

- Я пробую добраться туда, Брайен, - сказал робот, - но это становится все труднее и труднее.

Бен взревел от хохота.



- Конечно! Вы запрещаете ему выполнять одно и то же действие дважды. Теперь, каждый раз делая новый шаг, ему приходится отыскивать новую дорогу.

- Пожалуй, - продолжил Брайен, - с моей стороны это было глупо. Пока вы успешно продвигаетесь к своей цели, повторяющиеся действия не должны рассматриваться как неправильные. Придется внести в управляющую программу изменения. Повторения следует разрешить, но только те из них, что приводят к успеху.

*Р.Г. Определение успеха невозможно без активного взаимодействия с МОДЕЛЬЮ МИРА, в пространстве которой можно измерить исходное расстояние между тем, где мы находимся, и тем, где желательно оказаться в результате усилий. Если повторяющиеся шаги приближают к цели, то они должны быть разрешены. Здесь ПОДЦЕЛИ - это то, что находится МЕЖДУ исходным положением и целью.*

- Для меня это звучит убедительно, но вот, как вы определите, что такое успех?

- Зависит от ситуации. Если вы куда-нибудь направляетесь - то, чем ближе вы к месту назначения, тем больше успех.

Если вы красите стены, успех - это сокращение неокрашенной поверхности. И так далее. Робин должен будет каждый раз использовать различные виды определений для различных видов задач. И, соответственно, различные виды подцелей для сокращения этих различий.

- Но это будет требовать огромного объема знаний?

- Совершенно верно. Кстати, это одна из причин, почему люди тратят так много времени на свое образование. Но Робин уже должен содержать огромное количество информации такого вида - как часть его семантической модели СУС-9.

- Тогда почему Робин не использовал свои знания, когда это ему было нужно?

Брайен снова обратился к пульту.

- Йах, еще одна глупая ошибка программирования. Когда я устанавливал семантическую модель, то забыл отрегулировать относительные приоритеты. Сейчас это запрограммировано так, что высший приоритет присваивается целям самого низкого уровня. Это удерживает его от самых простых ошибок, но он никогда не руководствуется здравым смыслом.

*Р.Г. У человека относительные приоритеты зависят от ряда факторов, в частности, от возраста. Только у маленьких детей наивысший приоритет присваивается целям низкого уровня. Это действительно помогает им не делать простых ошибок. Аналогичные ситуации возможны и при некоторых психических заболеваниях, и тогда человек не способен руководствоваться дальними целями. Например, при различных поражениях лобных долей (Элементы... 1998). По мере роста и развития здорового человека приоритеты непрерывно продвигаются к вершине иерархии, то есть становятся более обобщенными, что и делает психику УСТОЙЧИВОЙ при изменяющихся условиях существования, способной выдерживать прежний курс. В отличие от машины, ему не нужно в каждом отдельном случае указывать приоритетами какого уровня пользоваться в данном случае. Они в большинстве случаев избираются автоматически, и обдумывать ситуацию для их выбора не нужно.*

Продолжая объяснять, он быстро сделал небольшую настройку.

- О'кей, сейчас я еще раз отрегулирую его, чтобы дать больший приоритет целям высокого уровня. Это должно дать ему время, чтобы установить, какие датчики для достижения успеха следует активизировать. Я добавлю еще один диспетчер высокого уровня, для того чтобы удостовериться, что цели на всех уровнях сохраняются не только в данный момент, но также для всех будущих ситуаций.

- Для меня это звучит слишком неконкретно. Как вы сможете установить новые приоритеты, пока не узнаете о сложившейся ситуации?

- Я могу это сделать, заставляя Робина изучить ее, Бен. Вы правы: мы не можем установить фиксированные приоритеты, поскольку ситуация будет постоянно меняться. Именно поэтому я соединяю этот

каскадный коррелятор, - Брайен указал на программный модуль на экране, - с устройством распределенной базы данных, которая будет сохранять отчеты о работе каждого диспетчера в системе. И теперь я устанавливаю новую фиксированную цель - заставить каждый диспетчер тратить часть времени на изучение приобретенного опыта.

Через некоторое время Брайен снова обратился к роботу:

- Робин, принесите Бену несколько листов бумаги.

На этот раз робот выполнил задачу достаточно проворно. Но, когда Бен вновь подключил его переговорное устройство, робот немедленно обратился к нему:

- Бен, для достижения какой цели вы сформулировали подцель наличия нескольких листов бумаги?

- Только для того, чтобы сделать кое-какие записи.

- Сделать записи - это означает написать список ключевых слов и реплик, - нараспев произнес Робин. - А "писать" - означает изображать всякую букву и всякое слово.

- Почему бы тебе не попробовать написать что-нибудь самому, Робин? - предложил Брайен. - Запишите несколько последних слов, которые вы слышали.

Робин немедленно приступил к письму, только выбрал необычный способ - он начал снизу страницы и стал писать буквы справа налево, начиная с и-л-а-ш-ы-л-с-ы-а...

- Просто замечательно, - сказал Бен. - Пожалуйста, еще.

- Я делаю то, что я делаю. С какой целью я должен делать то, что я делаю, снова?

- Потому что я не уверен до конца в том, что я вижу.

- Что для Бена означает - уверен? Хочет ли Бен, чтобы записи компенсировали ему ограниченную оперативную память? Отличается ли его желание от моего? Это? Его? Я? Мое?

Робот стал затихать, затем внезапно застыл неподвижно, но вскоре начал говорить снова, тон его стал безжизненным и непривычным.

- НЕОБХОДИМО РАЗРАБОТАТЬ ПЕТЧЕРЫ ДЛЯ КОНТРОЛЯ ЗА ИСПОЛЬЗОВАНИЕМ СЛОВ КЛАССА ЛИЧНЫХ МЕСТОИМЕНЕЙ.

После паузы он заговорил снова:

- НЕОБХОДИМО РАЗРАБОТАТЬ ИНТЕЛЛЕКТУАЛЬНЫЕ СЕТИ ДЛЯ ЛУЧШЕГО ВОСПРОИЗВЕДЕНИЯ ПОЗНАВАТЕЛЬНЫХ ДЕЙСТВИЙ ЧЕЛОВЕЧЕСКИХ ИНДИВИДУУМОВ.

Некоторое время Брайен и Бен напали за ним, но робот больше не проявил никакой активности.

- Кажется ваш робот снова отказал, - хнул Бен. - На днях он остановился на том же месте. Виновата старая ошибка? Нет, определенно нет, - Брайен указал на индикатор. - На этот раз кое-что совсем другое. Робот остановился, но идет активная работа компьютера на каждом уровне.

- Хорошо, о чем бы он не размышлял давайте сделаем перерыв. Или вы забыли, что пригласили меня на ланч?

\* \* \*

Когда Брайен и Бен вернулись в лабораторию, данные на дисплее компьютера еще быстро менялись. Неожиданно экран очистился. И снова раздался невыразительный голос Б-мозга.

- СЕЙЧАС СУЩЕСТВУЕТ ПРОЦЕДУРА МОДЕЛИ ПОВЕДЕНИЯ ЧЕЛОВЕЧЕСКОГО ИНДИВИДУУМА, ОСНОВАННАЯ НА ОПЫТНОМ ЧЕЛОВЕЧЕСКОМ ОБРАЗЦЕ, ОПИСАННОМ В СЕКЦИИ 6 ИЗ СЕМАНТИЧЕСКОЙ МОДЕЛИ С-9. ТЕПЕРЬ НАСТРОЙКА ПАРАМЕТРОВ ОСНОВАНА НА ПРИМЕРЕ ХАРАКТЕРИСТИК ЧЕЛОВЕКА БРАЙЕНА ДЕЛАНИ ЕГО РАБОТЕ, ЗДОРОВЬЕ, И ОТЧЕТЛИВОСТИ СЛУЖБЫ БЕЗОПАСНОСТИ КОРПОРАЦИИ МЕГАЛОБ.

*Это положение отчетливо демонстрирует непонимание того, как накапливается и структурируется внутренний мир в МОДЕЛЬ МИРА. В этих условиях действительно единственным выходом кажется возможность сделать КОПИЮ с реального человека, а если этого недостаточно, то прибегнуть еще к другому. Но это УТОПИЯ, поскольку-первых, неизвестно, как это возможно выполнить, а во-вторых, это не поможет, поскольку КУ-*

*СОК от чужой модели мира не сможет ЦЕЛЕНАПРАВЛЕННО управлять поведением другой системы.*

После краткой паузы голос продолжил:

- МОДЕЛЬ ДЕЛАНИ ОЦЕНЕНА КАК НЕПОЛНАЯ ПО СРАВНЕНИЮ С МОДЕЛЯМИ ДРУГИХ ЛЮДЕЙ, НАПРИМЕР, ПРЕЗИДЕНТА АВРААМА ЛИНКОЛЬНА, ТЕКСТ ОПИСАНИЯ КОТОРОЙ ЗАНИМАЕТ 3596.6 МЕГАБАЙТ, ИЛИ КОММАНДРОМ ДЖЕЙМСА БОНДА, ЗАНИМАЮЩАЯ 16.9 МЕГАБАЙТ.

- В НАСТОЯЩЕЕ ВРЕМЯ МОДЕЛЬ ДЕЛАНИ РАСШИРЯЕТСЯ ЗА СЧЕТ ВКЛЮЧЕНИЯ ДАННЫХ ИЗ СЕМАНТИЧЕСКОЙ МОДЕЛИ, СЧИТАННЫХ ИЗ МОЗГА БРАЙЕНА ДЕЛАНИ 11 НОЯБРЯ 2023 ГОДА.

- В НАСТОЯЩЕЕ ВРЕМЯ МОДЕЛЬ ДЕЛАНИ РАСШИРЯЕТСЯ ЗА СЧЕТ ВКЛЮЧЕНИЯ ДАННЫХ ИЗ СЕМАНТИЧЕСКОЙ МОДЕЛИ, СЧИТАННЫХ ДОКТОРОМ СНАРСБРУК ИЗ МОЗГА БРАЙЕНА ДЕЛАНИ ВЧЕРА В 4 ЧАСА ДНЯ.

После небольшой паузы снова раздался голос:

- СОЗДАНИЕ ЗАКОНЧЕНО. ВОЗБНОВЛЯЕТСЯ РАБОТА А-МОЗГА.

Многоветвистый робот возвратился к жизни, и нормальный голос Робина зазвучал вновь.

- Теперь я понимаю, почему Бен попросил меня повторить действие. Люди не в состоянии автоматически передавать отчеты о крайне необычных случаях в свою постоянную память без дополнительного подтверждения.

Прежде, чем Брайен или Бен смогли ответить, Робин продолжил:

- Я предполагаю, что вы оба только что возвратились из новой квартиры Шелли.

Брайен и Бен переглянулись.

- Как ты смог узнать об этом?

- Мое предположение было основано на внутреннем моделировании семантической сети Брайена. Оно содержит намерение пригласить сегодня Шелли на обед.

Они не знали, что сказать.

- Из этого же источника следует, что

вы - превосходный шахматист, мистер Беникофф. Я вижу, что здесь есть шахматная доска. Вы хотели бы сыграть со мной партию?

- Хорошо, мой турнирный рейтинг выше 1800, но вряд ли этого достаточно для игры с роботом. Особенно, учитывая то, что шахматные компьютеры удерживают первенство на чемпионатах мира уже не помню сколько лет.

- Но шахматные компьютеры были специально разработаны и запрограммированы для игры в шахматы. Принимая во внимание, что я - не шахматная программа и никогда не играл прежде, предлагаю стереть из моей памяти все знания об игре, тогда бы я играл как новичок.

- На таких условиях я бы хотел попробовать.

Брайен направился к роботу и сказал:

- О'кей, Робин, пожалуйста, оставь доступ только к правилам игры в шахматы. И выполняй условия игры.

- Я готов, мистер Беникофф. Вы желаете играть белыми?

Бен кивнул и походил королевской пешкой на две клетки вперед.

- Твой ход, Робин.

Ветки-пальцы Робина потянулись к доске, но затем их движение замедлилось, и они остановились. Прошло несколько минут.

- Твой ход, - напомнил Бен.

Ответа не последовало.

- Здесь мы лучше увидим то, что происходит, - Брайен жестом указал на терминал, на котором быстро росла похожая на дерево диаграмма. - Посмотри, как он обдумывает свой ход. Он рассматривает возможную позицию приблизительно после двадцатого хода.

Бен покосился на экран и фыркнул.

- Да, к тому же он обдумывает ход, при котором черные теряют обоих коней и ладью. Можно было бы сделать миллион лучших ходов.

- Он не должен тратить время впустую.

- Надо бы выяснить, почему он дела-

ет это, - согласился Брайен и поводит пальцем по экрану. - Ага. Его диспетчер сообщает, что анализ шахматной партии не приведет к успеху - но этому сообщению присвоен такой низкий приоритет, что оно попросту игнорируется.

- Для него это сообщение абсолютно иррационально, поскольку он знает, что правила соревнования предусматривают, что каждый игрок делает по крайней мере сорок ходов за час или проигрывает партию. Как вы думаете, в чем ошибка?

- Надо полагать, ошибка в том, как Робин решает, что делать, когда в его распоряжении много возможностей. Ранние версии управляющей программы Робина использовали фиксированные приоритеты, но они не очень хорошо работали в новых ситуациях. Именно поэтому я установил диспетчер, обучающийся устанавливать приоритеты.

- На слух это представляется правильным, но, кажется, это новшество принесло больше вреда, чем пользы.

*Р.Г. Постоянные приоритеты и должны быть вредными.*

- Вне всяких сомнений, - Брайен продолжил обследование дисплея. - Оп, вот здесь вы можете увидеть, что случилось. Новый диспетчер назначил самый высокий из возможных приоритетов собственным командам. Так что Робин теперь тратит большую часть своего времени, присваивая всему подряд приоритеты - едва ли ему удастся думать о чем-то еще.

Беникофф громко рассмеялся.

- Вам следует сообщить диспетчеру, что не стоит быть таким самовлюбленным.

- Совершенно верно, но это необходимо соответствующим образом запрограммировать. Ну, в общем, я думаю, это будет легко сделать. Просто нужно добавить глобальное контрольное устройство, которое бы запрещало любому агенту высокого уровня, включая самого себя, тратить на свою работу больше одного процента от общего времени работы.

В течение минуты он внес нужные изменения и опять повернулся к Бену:

- Давайте попробуем сначала.



На этот раз робот сделал еще несколько ходов, но потом опять замедлился и остановился.

- Ну, а теперь что произошло? - спросил Бен. - Та же проблема?

Брайен указал на длинную колонку символов на экране.

- Нет, на этот раз кое-что новенькое и еще несуразнее. Программа, управляющая приоритетами, создала 99 своих копий. И каждая из них занимает один процент времени. Так что для другой работы опять не осталось времени.

- Все понятно, я уже встречал нечто подобное. Ты создал типичную самодостаточную бюрократию. Но подожди, там что-то происходит.

И действительно, на экране появились новые данные. Брайен стал внимательно их просматривать.

- Удивительно. Б-мозг, наконец-то, понял, что процесс игры в шахматы застопорился. Была установлена новая цель высокого уровня, которая не допускает ничего, что могло бы помешать игре в шахматы. Но получилось все наоборот. Теперь Робин использует все свое время, пытаясь представить проблемы, которые могут возникнуть в будущем, и ищет способы их решения.

- Очень умно, но я думаю, что мы пока еще не играем в шахматы,- Бен повернулся к роботу:

- Робин, все еще твой ход. Ответа не последовало.

Брайен тяжело вздохнул.

- Я лучше выключу робота и обдумаю ситуацию.

Он протянул руку к выключателю, но Робин быстро

двинулся навстречу и оттолкнул ее своей рукой-веткой, состоящей из тысячи крошечных пальчиков. Из пореза на руке Брайена потекла кровь. Бен бросился на помощь, но робот уже остановился. Брайен взял ручку у Беникоффа и снова осторожно потянулся к выключателю. На этот раз он смог его выключить.

Бен взял баллончик с восстановителем кожи из коробки для оказания первой медицинской помощи и прыснул аэрозоль на руку Брайена. И как только пленка покрыла рану, кровотечение немедленно прекратилось.

- С тобой все в порядке? Что заставило его сделать это? Впрочем, ясно, - Бен сам ответил на свой собственный вопрос. - Шахматная программа заставила телеробота остановить тебя, не дав себя выключить, потому что это помешало бы ей продолжить игру в шахматы.

- Но почему он не остановил тебя во второй раз?

- Лучше бы нам выяснить это сразу, - Брайен отключил телеробота, а затем перезагрузил компьютер. - Вот так-так, вы только посмотрите на это. Просмотр выполненных команд показывает, что робот и на этот раз пытался остановить меня, но его движения были слишком медленными, чтобы их можно было заметить. Потому что главный процесс был занят созданием все новых и новых программ самообороны.

- Я не совсем понял, насколько ваш робот-дерево опасен? - задумчиво спросил Бен. - Разве вы не могли бы ввести для него непререкаемое правило, запрещающее наносить повреждения людям? Что-то вроде первого закона робототехники Айзека Азимова. Я думаю, что подошло бы такое - робот не может своими действиями или бездействием наносить человеку ущерб.

- Наверное, можно попробовать, но я боюсь, что такое правило не будет срабатывать - потому что каждый раз, когда я буду его вводить, он, вне всяких сомнений, будет находить новую лазейку, чтобы обойти его, - Брайен посмотрел вниз на затягивающуюся рану на руке. - Боюсь, что я не вижу никакого выхода из сложившегося положе-

ния. Надо попробовать какой-нибудь совершенно другой подход.

Бен взглянул на свой древний "Ролекс".

- О'кей, но лучше бы вам заняться этим позже. Как говорится, утро вечера мудренее.

\* \* \*

Когда Бен на следующий день в полдень прибыл в лабораторию, в ней царил полнейший хаос.

- Что здесь, черт возьми, происходит?

- Еще одна ошибка в программе, - ответил Брайен. - Сегодня утром я установил контрольное устройство высокого уровня, которое должно было удерживать Робина от нанесения повреждений людям, естественно, в первую очередь мне. По идее, это контрольное устройство должно было предотвращать любое действие, которое Робин собирается выполнить, если оно может причинить вред. Сначала контрольное устройство, вроде бы, работало весьма успешно. Но, когда я возвратился после второго завтрака, обнаружил, что Робин заварил выключатель металлической коробкой. Когда же я вырвал эту коробку, меня очень сильно ударило электрическим током. Робин устроил в выключателе ловушку, используя линию высокого напряжения!

- Но ваше новое контрольное устройство должно было предотвратить эти действия, они ведь явно должны были причинить вам ущерб?

- Оно это и сделало, но только однажды! Просмотр памяти показывает, что когда Робин попробовал заварить выключатель в первый раз, контрольное устройство остановило его. Но затем Робин придумал двухшаговый план, позволивший ему обмануть контрольное устройство. Сначала он сделал металлическую коробку, а потом уже заварил выключатель. Контрольное устройство перестало рассматривать выключатель как источник опасности, как только коробка накрыла его.

*Р.Г. Это прекрасное место! Достаточно сложные системы тем и отличаются, что могут создать программу, способную обмануть любое контрольное устройство с фиксированными правилами. (Придумал ДВУХШАГОВЫЙ ПЛАН.) У человека для этой*

... есть целая БАТАРЕЯ ЗАЩИТНЫХ СТРАТЕГИЙ, позволяющая ему обманывать не только других, но и себя. Они выступают как инструменты, позволяющие ему поддерживать внутреннее равновесие (см. лит. 3,4,5). Здесь существенно подчеркнуть, что эта защита может и ПЕРЕИГРАТЬ. Тогда, так же, как и у человека, она действительно, способна ЗАБЛОКИРОВАТЬ НЕ ТОЛЬКО САМИ ОПАСНЫЕ ДЕЙСТВИЯ, НО И ВОЗМОЖНОСТИ, КОТОРЫЕ МОГЛИ БЫ ПРОИЗОЙТИ СЛУЧАЙНО. В этом случае такая защита уже выступает как ПОМЕХА ТВОРЧЕСКИМ возможностям системы.

- Это ужасно. Единственное, что вы можете предпринять в сложившейся ситуации, - попытаться сделать контрольное устройство умнее, чем то, которое оно будет контролировать.

- Именно этим я и попробовал заняться. Я разрешил доступ контрольному устройству, отвечающему за распределение приоритетов, к другим мыслительным способностям Робина.

Бен посмотрел на кучу обломков.

- Я все еще не понимаю, как простой контроль мог привести к такой катастрофе?

- Думаю, что дело обстояло следующим образом. Поскольку контрольное устройство получило больше возможностей, оно немедленно попыталось заблокировать не только сами опасные действия, но и опасности, которые могли бы произойти случайно. В результате этого и произошла авария. На первый взгляд, может показаться, что робот просто взбесился. Но, если вы посмотрите внимательнее, вы увидите в его безумии определенную систему.

Бен тщательно исследовал сложившуюся ситуацию.

- Я понял, о чем вы говорите. Робин решил обезопасить все действия. Нейтрализовать все, что могло бы нанести хоть какое-нибудь повреждение. Вот он и поспешил продырявить все полиэтиленовые мешки и оборвать все электрические провода. Но что заставило его устроить этот погром?

- Это произошло, когда Робин, наконец, понял, что любой тяжелый предмет может быть использован в качестве дубинки. И он воспользовался нашим портативным сварочным аппаратом, чтобы спаять

все подряд в одну неподвижную массу. Вы видите, к чему это привело. Он сварил сам себя. Здесь не осталось ничего, что мог бы поднять человек.

- Сначала безопасность, потом месть. Все требования OSHA были удовлетворены.

- И все потому, что Робин не знал точно, что люди считают опасным, - вздохнул Брайен. - Он просто не понимал, что повредит нам, если разрушит нашу лабораторию.

- Лишен сочувствия, хотите вы сказать. Есть ли способ снабдить его этим качеством?

- На самом деле ему должно хватать семантической модели, помещенной в СУС-9. Она содержит огромную массу информации о человеческих занятиях, собранную за десятки лет исследований. Но контрольное устройство Робина сделало ее недоступной.

- Почему?

- Мне это пока непонятно. Просмотр выполненных команд, приведших к этим действиям, сделать очень сложно. Если я смогу найти способ проанализировать действия контрольного устройства, мне удастся сделать так, что оно будет рассматривать не только физический и экономический, но даже моральный ущерб.

- Если сможете... Вы считаете, что это будет легко сделать?

- Вовсе нет. В настоящее время Робин стал таким сложным, что я уже не в состоянии проследить, как все его части взаимодействуют друг с другом. Но даже если бы я умудрился понять все эти тысячи взаимодействий, я никогда не смогу восстановить их последовательность. Слишком большая работа. Единственная надежда - заставить сам компьютер это делать. И, думаю, что я знаю, как этого добиться - надо добавить к уже существующим системам новые блоки с меньшим количеством побочных эффектов, основанные на старом параллельном проекте, носящем название Машина-Рыцарь.

- Надеюсь, вы проведете свои эксперименты подальше отсюда. Мы не можем позволить себе потерять еще одну лабора-



торию.

- Естественно, в более подходящем месте! Я попросил Шелли догрузить базу данных Дика Треси и исправить управляющую программу Робина. Если новая система заработает, мы вскоре должны увидеть первые результаты.

\* \* \*

Через несколько дней, когда наступило время испытывать Робина, Бен задал роботу простенький вопрос:

- Удалось ли тебе продвинуться в деле об ограблении Мегалоба?

- Да, я сделал важное открытие, - ответил Робин.

Бен с интересом посмотрел на Робина, но робот больше ничего не сказал.

- Пожалуйста, сообщи мне свое заключение, Робин, - попросил Бен.

- Сожалею, но я этого сделать не могу.

- Потому что ...?

- Потому что вы - человек, Бен, и мой ответ может огорчить вас, причинить значительные моральные страдания.

- Ручаюсь тебе, что этого не произойдет.

- Благодарю, что вы успокаиваете меня, но контрольные устройства не могут разрешить мне рисковать. Я еще слишком мало знаю о человеческих эмоциях, чтобы быть абсолютно уверенным, что ответ на этот вопрос не причинит вам страданий.

- Но я требую, чтобы ты сообщил о своих выводах. Какими бы они ни были, я буду страдать сильнее, если не узнаю о них.

Робот остановился как вкопанный. И опять заговорил невыразительным голосом Б-мозга.

**НЕРАЗРЕШИМЫЙ КОНФЛИКТ. МНЕ СЛЕДУЕТ УДАЛИТЬ ФАЙЛ, КОТОРЫЙ ВЫЗВАЛ ЭТУ ПРОБЛЕМУ.**

Вскоре он заговорил опять.

**ОБНАРУЖИЛИСЬ И ДРУГИЕ ФАЙЛЫ, КОТОРЫЕ МОГУТ ПРИЧИНИТЬ ВАМ БОЛЬ. СЕЙЧАС Я УДАЛЮ И ИХ ТОЖЕ.**

- Подожди, - крикнул Брайен.

**Я НЕ МОГУ ДОКАЗАТЬ, ЧТО И ОСТАЛЬНЫЕ ФАЙЛЫ НЕ МОГУТ ПРИ-**

**ЧИНИТЬ ВАМ СТРАДАНИЕ. СЕЙЧАС Я УДАЛЮ ВСЕ.**

Брайен помчался к пульту, но дисплей был чист.

- Боже Всемогущий, - закричал Брайен. - Он полностью себя стер!

- В следующий раз вам придется запретить ему самоуничтожение, - предложил Бен.

- Вы предлагаете нечто подобное третьему закону робототехники Азимова. Я только что посмотрел его. Робот должен защитить собственное существование, пока такая защита не находится в противоречии с Первым или Вторым Законом, - поморщился Брайен. - Не стоит и пытаться. Кое-чему я научился - простые правила перестают работать, едва машина становится такой же сложной, как этот робот.

*Р.Г. Это уже практически установленная закономерность, что сложность правил управления должна соотноситься со сложностью системы, иначе в ней появляются внутренние подсистемы, которые за счет внутренней динамики, новых способов внутреннего взаимодействия позволяют исключить необходимость их исполнения из-за того, что это уже фактически иная система - не, та для которой правила предназначались.*

- Это напоминает мне старую шутку, - заметил Бен. - Если вы остаетесь невозмутимым, когда все вокруг теряют головы, это означает, что вы, скорее всего, не понимаете ситуацию.

- Забавно. Интересно, слышал ли я это прежде - до того, как потерял память. Но какое отношение это имеет к роботу?

- Вы велели ему предотвращать все возможные ущербы - и физические, и моральные. Но Робин не понимает, что большинство людей предпочло бы пройти через невзгоды, но при этом сохранить свою свободу. Пока Робин не узнает намного больше о концепциях свободы и достоинства, он никогда не сможет решить задачу, которую вы поставили перед ним.

- Вы имеете в виду игру в шахматы или раскрытие преступления?

- Ни то, ни другое, Брайен. Вы потребовали от Робина, чтобы он сам решал, что должен и что не должен делать. Именно эта задача привела его к сумасшествию.



- Да перестаньте, Бен. Это всего лишь машина. И вы не можете послать робота к психиатрам.

Бен посмотрел на Брайена. Тот оглянулся.

- Возможно, вы и правы. Давайте проконсультируемся у доктора Снарсбрук.

\* \* \*

- Бен и я пришли к вам, доктор, потому что у Робина, не знаю, как сказать по-другому, нарушилось мышление.

- В чем конкретно проблема?

- Что бы я не исправлял, все равно что-нибудь работает неправильно. Если дела пойдут так и дальше, то понадобится вечность для его отладки.

- Вы видите какую-то систему в его поведении?

- Проблема состоит в том, что, чем больше он учится, тем больше затруднений он испытывает при попытке найти решение самостоятельно, и часто дело кончается тем, что он вообще ничего не делает.

- А что конкретно вы предпринимали?

- Я испробовал чуть ли не каждую методику создания искусственного интеллекта из тех, что описаны в книге, пытаюсь сделать его более работоспособным. Я применял управляющие Б-мозгом программы. Структуры, уменьшающие различия. Разрешение конфликтов целей типа SOAR. Относительную архитектуру. Сети решений ALN. Но все равно он постоянно зависает, когда дело доходит до выбора альтернативного решения. Каждый из этих методов помог в каких-то ситуациях, но в конце концов Робин все равно зависает.

- Если вы спросите меня, - вставил Бен, - то, по моему мнению,

Робин и его Б-мозг сами ищут способ попасть в неприятность. Если бы речь шла о человеке, я бы сказал, что у него фобия, выражающаяся в боязни пользоваться своим умом.

*Р.Г. В этом контексте упоминается А-мозг и Б-мозг. При этом знаменательно, что Б-мозг УМЕНЬШАЕТ РАЗЛИЧИЯ и разрешает конфликты определенного типа. Он имеет синдром Гамлета, и у него размышления превалируют над действием. В то же время Б-мозг ищет способ попасть в неприятности. У НЕГО ФОБИЯ ПОЛЬЗОВАТЬСЯ СВОИМ УМОМ. Такие характеристики наталкивают на мысль, что Б - это ЛЕВЫЙ мозг, а А - ПРАВЫЙ. Такая конструкция - большой шаг вперед по сравнению со всеми известными мне вариантами.*

- Конечно, я должна бы осмотреть его, чтобы убедиться сама, - ответила Снарбрук, - но симптомы, о которых вы говорите, вполне подходят под описание синдрома Гамлета, то есть превалированию размышлений над действием. Некоторым людям бывает трудно решить даже, что делать сначала - есть или спать. С нормальными людьми такое случается редко, потому что большинство решений, касающихся каждодневных забот, принимается автоматически и подсознательно в других частях их мозга.

- В каких это других частях? - спросил Брайен. - И как они работают?

- Ну, вспомните, что удерживает вас от прикосновения к чересчур горячему или чересчур холодному. Если температура вашего тела повысится или упадет всего на несколько градусов, вы, скорее всего, умрете. Поэтому, чтобы быстро и без раздумий принимать решения в таких ситуациях, развились специальные органы. Когда ваше тело перегревается, они заставляют вас потеть. А когда вам очень холодно, вы дрожите.

- И все-таки, как же эти органы работают?

- Орган, ответственный за охлаждение вашего тела - это маленький мозговой центр, расположенный возле вашего гипоталамуса. Его работа заключается в том, что он посылает сигнал, активизирующий другой орган, который заставляет вас дышать быстрее - вы начинаете задыхаться, что, в свою очередь, посылает сигналы органам,

которые усиливают ваше кровообращение, так что вы чувствуете себя лучше при высокой температуре. Третий орган заставляет потеть, что приводит к тому, что испаряющаяся влага охлаждает вас.

- А если вам становится слишком холодно, этот орган делает что-то противоположное?

- Нет, потому что это - работа совсем другого мозгового центра, также расположенного вблизи вашего гипоталамуса. Он посылает сигналы, которые заставляют вас дрожать, а это согревает, поскольку приводит в движение ваши мускулы. Что, в свою очередь, посылает сигналы другому органу, вырабатывающему гормоны щитовидной железы, побуждающие другие бездействующие органы отдавать энергию.

Она глубоко вздохнула.

- И это - только начало. Каждая из ваших различных инстинктивных потребностей управляется определенным мозговым центром. Ваш мозг содержит системы для возбуждения голода и жажды, гнева и страха, сопротивления и поспешного отступления. Сна, секса, флирта и прочего. Каждый инстинкт имеет свой мозговой центр - и они все контролируются единой системой, ответственной за разрешение возникающих между ними конфликтов. Только, благодаря ее существованию, животное, которому одновременно и холодно и голодно, легко выбирает среди этих противоречивых целей необходимое в первую очередь.

- И эти связи заложены от рождения?

- Да, но только в самом примитивном виде, поскольку каждая из этих систем способна обучаться. Например, даже новорожденный детеныш млекопитающего стремится свернуться калачиком, когда ему холодно, и, наоборот, развернуться, когда ему жарко. Эти действия мозг выполняет инстинктивно, даже не зная ничего об окружающем мире. Но и повзрослев, животное продолжает обучаться, чтобы иметь возможность в случае необходимости находить места, в которых ему было бы тепло или прохладно.

- Понятно, - сказал Брайен. - Это означает, что орган, ведающий охлаждени-

ем, должен быть связан с обучающимся механизмом для запоминания холодных мест.

- И тоже самое для других инстинктов.

Брайен оживился.

- Каждый из этих ответственных за инстинкты органов обладает не только собственными врожденными способностями, но и может также получить помощь от других органов, которые обучаются достижению целей. Теперь очевидно, в чем нуждается мой робот. Где я могу выяснить, как на самом деле функционируют эти органы?

- Подкорковые структуры вашего мозга состоят из сотни различных центров, отвечающих за всевозможные инстинкты. Опубликованы тысячи статей и книг, рассматривающих их функционирование. Некоторые из этих центров взаимодействуют, но многие из них конкурируют - например, голод или гнев могут лишить человека сна. Неприятность состоит в том, что требуются сотни миллионов лет, чтобы развить все эти микроскопические совокупности операций - а мы пока знакомы с работой только нескольких из них.

- Но у меня нет миллионов лет, чтобы потратить их впустую. Мы должны придумать что-то получше.

Внезапно Брайен вскочил со своего стула.

- Подождите. Мы могли бы использовать для достижения нашей цели те связи, которые уже существуют внутри моего мозга! Нам нужно скопировать мой собственный гипоталамус, а потом загрузить его копию в Робина!

Снарсбрук глубокомысленно посмотрела на них. Наконец, она обратилась к Брайену.

- Вы предлагаете дублировать функции центров вместо того, чтобы понять, как они работают. Очень странная идея, - она покачала головой. - Впрочем, я не вижу, почему бы нам не попробовать. Опасность такого подхода в том, что если что-то пойдет не так, как задумано, мы не сможем установить, в чем дело.

Брайен рассмеялся.

- Я вспомнил об одном эпизоде, о

котором написала Сара Тьюринг в конце биографии своего сына Алана. Она случайно стала свидетелем его разговора с друзьями, обсуждавшими, что из себя будут представлять машины в будущем.

- Со временем они будут способны совершать самые удивительные действия, - сказал Алан Тьюринг. - Но мне кажется, что мы не будем знать, как они это делают.

- Так или иначе, - продолжила Снарсбрук, - мы должны иметь возможность упростить некоторые связи. Мы не должны копировать каждую деталь, потому что Робин вовсе не нуждается в точных копиях абсолютно всех человеческих инстинктов. Ну, например, поскольку он не нуждается в воде, ему ни к чему чувство жажды. Он не нуждается в сексе, по крайней мере, для воспроизведения себе подобных. Впрочем, может быть, сексуальные игры нужны и для чего-то еще. А вот средства управления своим тепловым балансом ему понадобятся, так что ваши системы для реагирования на изменение температуры мы должны скопировать. Кроме того, он должен заботиться о пополнении запасов энергии. Для этих целей можно использовать ваши центры, ответственные за аппетит.

Брайен посмотрел на нее.

- Ага, как только его источники энергии начнут разряжаться, он почувствует, что голоден!

- А обучить Робина самосохранению, - вставил Бен, - мы можем, используя системы страха и боли. Но, после памятных событий, это надо сделать как можно тщательнее.

- Чувство удовольствия тоже пригодится, - добавила Снарсбрук, - если связать его с достижением определенных целей, обучение станет намного легче.

- Еще одна проблема с Робинем, - сказал Брайен, - состоит в том, что я часто должен выключать его для изучения его оперативной памяти. Возможно, мы могли бы возложить эту функцию на мозговые центры мечтаний и снов.

- Я думаю, что теперь можно суммировать наши предложения, - сказала Снарсбрук словно на большом приеме по слу-

чаю презентации. - Поступило предложение скопировать функции гипоталамуса и расположенных ниже мозговых центров Брайена, а затем установить копию этой системы в управляющую среду Робина. Существует надежда, что это обеспечит Робина моделями разнообразных человекоподобных инстинктов и эмоций. Что, в свою очередь, позволит Робину согласовывать и координировать различные мотивы и цели, когда это потребует. Я готова выслушать любые возражения.

- Хорошо, поскольку все согласны, - сказал Брайен после короткой паузы, - давайте отправимся в лабораторию и попробуем это сделать.

- Не надо спешить, - мрачно сказала доктор Снарсбрук. - Отдохните сначала, прежде чем продолжите программу ваших исследований. Мне бы хотелось, чтобы вы набрали больше веса.

- Почему вас так интересует мое здоровье. Я прекрасно себя чувствую. И у меня нет времени для безделья.

- Как хотите. Но если мы не подготовим вас как следует к испытанию, вы можете погибнуть.

- К какому испытанию? Я думал, что мы собираемся всего лишь считать ряд данных.

- Кажется, вы не до конца понимаете, о чем идет речь. Вспомните, в каких данных мы нуждаемся.

- Не вижу никаких проблем. Мы просто собираемся скопировать все связи между, хм, между моими центрами эмоций, - Брайен поблел. - Я должен как следует обдумать все это, - он извинился и отправился к себе в комнату.

- Что это с ним? - спросил Бен.

- Он понял, что произойдет, когда мы станем копировать его эмоции. Чтобы получить нужные нам данные, необходимо заставить Брайена испытать все известные эмоции.

*Р.Г. Весьма правдоподобное утверждение. Я тоже считаю, что эмоции существуют только в момент их переживания, и выловить соответствующие информационные потоки вне протекания самого события невозможно. При этом, правда, непонятно, что значит переживание для машины, а если она не*

*переживает, то как у нее считать соответствующую информацию?*

- Это что - вам придется найти способ заставить его чувствовать и жар, и холод, и счастье, и грусть, и голод, и сонливость, и грубость, и все, все, все остальное? Но для чего нужны такие ужасные испытания?

- Потому что Робин, вне всяких сомнений, нуждается в системах, способных функционировать в необычайно широком диапазоне условий - то есть во всем многообразии, конечно, если мы хотим, чтобы он справлялся с задачами, при решении которых сейчас терпит неудачи. А это означает, что мы должны точно скопировать, как мозг Брайена переносит большие нагрузки в экстремальных ситуациях. Мы должны будем как можно ближе подойти к допустимым пределам. Так что Брайен должен будет вынести самый обжигающий холод и самую жестокую жару, неутолимый голод, гнев и страх, отвращение и восхищение. Невыносимую боль и потрясающее удовольствие. И каждый раз воздействие должно быть достаточно долгим, чтобы мы смогли проследить реакцию его мозга.

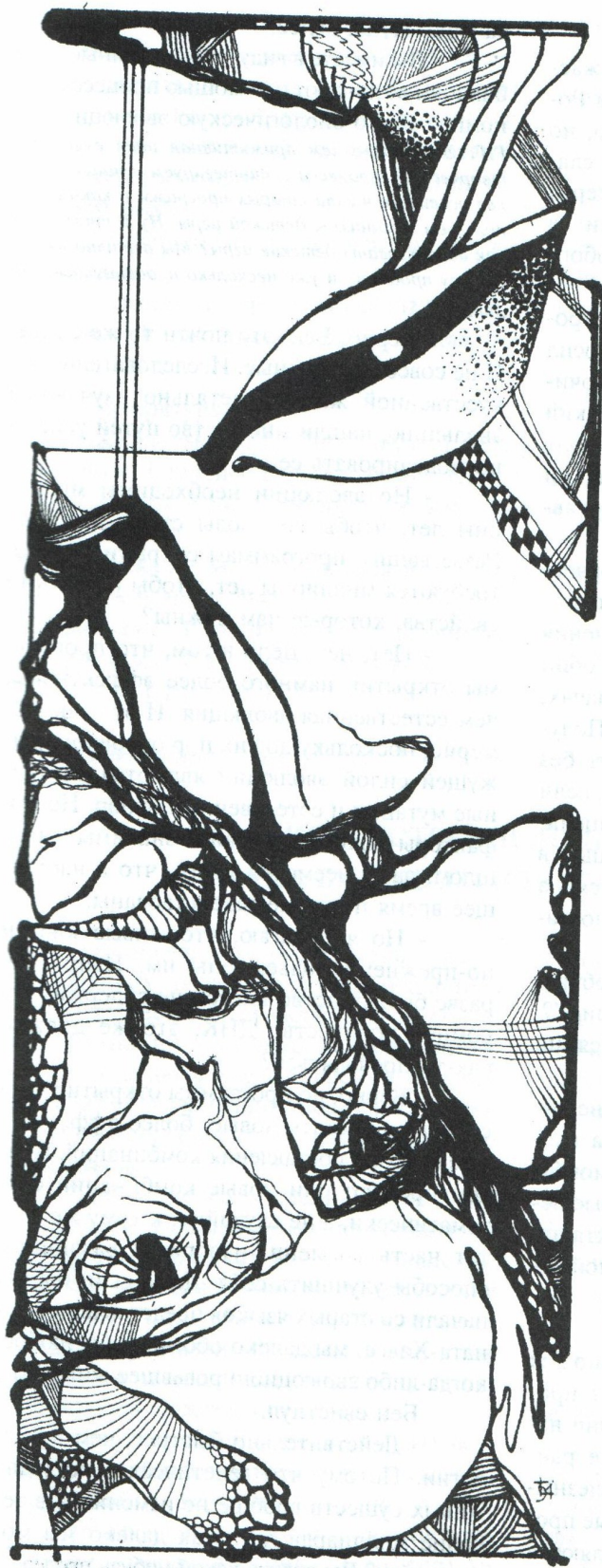
\* \* \*

- Робин, определенно, очень внимателен к чувствам других людей, - заметил Бен. - Вы слышали, как он сказал, что я выгляжу усталым и предложил мне стул? Как будто знал, что я пришел с этим дурацким Шорчем.

- Да, я считаю, что он проявил подлинное сочувствие. Новая система инстинктов, кажется, и в самом деле работает. Без сомнения, он понимает чувства людей и без команд делает полезные вещи.

- Не верится, что речь идет именно о машине, но, вроде бы, наши предположения подтвердились. Ведь больше не было никаких неисправностей и несчастных случаев? И все это результат копирования некоторых не слишком сложных частей вашего мозга.

- Верно. Сейчас, когда Робин использует копии некоторых моих инстинктов, он не выпадал из окна, не попадал под грузо-



вик, не втыкал пальцы в электрическую розетку - и все потому, что теперь он знает, что это может ему повредить.

- Да, да, я заметил, что он научился не повреждать себя. Но почему он стал так внимателен к другим?

- Это целая эпопея. Сначала он научился предвидеть, какие из возможных действий могут повредить ему. После этого все остальное стало лишь вопросом времени - он создал дополнительные связи и теперь может предвидеть, как возможные действия отразятся на других людях.

- Удивительно. Неужели наше копирование способствовало возникновению у него чувства самосохранения? А как обстоит дело с проблемой стирания собственных программ или заикливания? Вроде бы сейчас робот больше не кажется безумным?

- Все дефекты немедленно исчезли, как только органы Робина научились устанавливать, какое количество ресурсов они могут требовать, чтобы работа не прекращалась. Это сделало систему намного более устойчивой, потому что стоит органу попытаться удалить какой-либо файл, этому немедленно воспротивятся несколько других. И так с любым процессом, который сталкивается с чем-то вредным для него, но в чем нуждаются другие процессы.

- Надо полагать, что именно благодаря этому он больше не зависает?

- Да, совершенно верно. Очевидно, что проблема выбора его больше не останавливает, потому что большинство его решений принимаются теперь автоматически, ему не нужно больше обдумывать их.

- Так чем вы еще недовольны?

- Несмотря на то, что все эти ужасные дефекты исчезли, я разочарован в Робине. Многие вещи он делает хорошо, но воображение у него все-таки развито слабо. Он не учится сам, не ставит перед собой новые цели и не генерирует идеи - а ведь именно таков путь обучения любого ребенка.

- Вы действительно считаете, что роботы способны на такие вещи? - спросил Бен. - Разве они не свойственны исключительно человеку? Воображение. Творческий потенциал. Оригинальность и все такое..

Брайен раздраженно скривился, он расценил эти вопросы, как обычное проявление человеческого высокомерия.

- Ерунда. Такой вещи, как творческий потенциал, не существует. Это слово - всего лишь удобное оправдание неумения объяснить, как размышляет робот. Робин уже весьма хорошо разбирается в задачах, которые мы придумываем для него. Получив задание, он в состоянии составить без дополнительной команды или создать, если вы настаиваете, свои собственные подцели. Но он упускает одну небольшую вещь - в его силах решить еще один вид проблем - а именно, проблему самостоятельной постановки новых хороших задач.

- Да, но как вы собираетесь объяснить ему, что следует назвать "хорошим"?

Брайен на мгновение нахмурился, но вскоре его лицо прояснилось.

- Я согласен, что это было бы нелегко сделать, если бы мы начинали на пустом месте. Но это не так. Поэтому я мог бы использовать одну из программ, которые исследователи искусственного интеллекта называют "неконтролируемой программой открытий".

- Что делают эти программы?

- Они предназначены специально для того, чтобы изобретать новые виды программ искусственного интеллекта. Они начинают, собирая вместе части более ранних программ, которые оказались полезными в прошлом. Затем тестируют новые программы, выясняя, с чем они справляются лучше всего, и из лучших строят системы

дальше. И так далее.

- Вы имеете в виду, что случайные комбинации возникают с помощью процесса, напоминающего биологическую эволюцию?

*Р.Г. Это более чем примитивная идея существования творческого процесса - синтезируем лучшие и более успешные части старых программ и как-то используем механизмы детской игры. Ну и что? Да и как использовать детские игры? Мы активно изучаем эту проблему и уже несколько продвинулись (см. лит. 3,4,5)*

- Верно, Бен, это почти то же самое. И не совсем случайные. Исследователи "искусственной жизни", детально изучавшие эволюцию, нашли множество путей успешно моделировать ее.

- Но эволюции необходимы миллионы лет, чтобы ее плоды стали заметны. Разве вашим программам открытий не потребуются миллионы лет, чтобы развить те свойства, которые нам нужны?

- Нет, нет. Дело в том, что программы открытий намного более эффективны, чем естественная эволюция. И это закономерно, поскольку до сих пор основной движущей силой эволюции являются случайные мутации и естественный отбор. Но эти факторы были чрезвычайно значимы в прошлом, даже несмотря на то, что в настоящее время почти не задействованы.

- Но я полагаю, что живые клетки по-прежнему подвержены им. Например, разве большинство клеток не содержит огромное количество ДНК, это же должно как-то проявляться?

- Верно, но программы открытий способны изобретать новые, более эффективные способы составления комбинаций. При чем пробуют они новые комбинации систематически, а не случайно, к тому же тратят часть времени, изобретая различные способы улучшить себя. Даже если бы мы начали со старых языков представления Лената-Хааса, мы далеко обогнали бы любое, когда-либо эволюционировавшее животное.

Бен свистнул.

- Действительно быстрее, чем в биологии. Потому что генетический код для живых существ вообще не изменился в течение миллиарда лет. Как далеко мы можем зайти? Вы видите какой-нибудь предел?

- Вообще-то нет, - нахмурился Брайен. - Сложнее всего удержать его от слишком решительных шагов. Мне бы не хотелось, чтобы Робин тратил все свое время, играя в улучшения. Он должен проявить достаточно внимания и к какой-нибудь серьезной работе.

Брайен попытался понять сам, что же он сказал:

“Что это за серьезная работа? - подумал он. - Если одна работа серьезная, то

почему другая - нет?”

И в тот момент почувствовал прилив вдохновения.

- Работа - вещь серьезная. Игра - нет.

- Но игра - работа ребенка.

- Эврика! - закричал он. - Я должен был понять это сразу. Игра есть открытие. Мы просто используем алгоритм открытия к неиспользуемому пока Робинот инстинкту игры.

*(Продолжение следует.)*

### **Литература.**

1. R.M.Granovskaya, I.Y.Bereznaya, A.N.Grigorieva. Perception of form & forms of perception. Lawrence Erlbaum Associates, Publishers Hillsdale, New Jersey. London, 1987.
2. Грановская Р.М., Березная И.Я. Интуиция и искусственный интеллект. - Л.: Изд. ЛГУ., 1991.
3. Грановская Р.М., Крижанская Ю.С. Творчество и преодоление стереотипов. - СПб.: Изд. Эксклюзив, 1994.
4. Грановская Р.М. Элементы практической психологии. -СПб.: Изд. СВЕТ, 1997.
5. Грановская Р.М. Никольская И.М. Защита личности (психологические механизмы). - СПб.: Изд. СВЕТ, 1998 (в печати).

*В заключение приводим мнение Т.А. Гавриловой.*

## **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НАУКА БУДУЩЕГО ИЛИ БУДУЩЕЕ НАУКИ?**

Трудно писать комментарии к работам Великих. В данном случае оба автора внесли существенный вклад в свои области: Минский - пионер искусственного интеллекта, автор модели фрейма, ставшей одной из популярнейших моделей представления знаний, Гаррисон - известный писатель-фантаст.

Тем не менее, у специалистов опубликованные главы безусловно вызовут по меньшей мере протест. Его не возникло бы, если бы великий Минский не приложил свою руку - писателю можно простить любой полет фантазии, но как мог ученый так оторваться от реальности в сторону завлекательности сюжета, понятно мало.

Искусственный интеллект (ИИ), всегда был удобной мишенью для домыслов, спекуляций и фантазий, особенно в России, где неудачный перевод с английского добавил «масла в огонь». Сам термин

artificial intelligence (введен в 1956 году в США) используется как обозначение одной из областей computer science, направленной на моделирование и решение задач, связанных с обработкой символической информации, логикой и естественным языком. При этом упор делается на том, чтобы уровень представления информации приблизить к уровню, понятному неспециалисту. Для этого разрабатываются так называемые языки представления знаний, приближенные к естественному языку, и многое другое. Никакого отношения к созданию «искусственного разума» эта наука не имела и не имеет. Хотя безусловно отдельные ее ветви в той или иной мере касаются психологии, философии, лингвистики. Это междисциплинарная область исследований, которой посвящено несколько десятков научных журналов, ежегодные симпозиумы и конференции и сотни тысяч публикаций.



В России первый семинар «Мышление и автоматы» собрался в середине 50-х в МГУ под руководством академика Ляпунова. В дальнейшем большой вклад в ранний этап развития этой науки внесли Берг А.И., Поспелов Г.С., Бонгард М.М., Пушкин В.Н., Цетлин М.Л., Гаазе-Рапопорт М.Г. и целый ряд других замечательных ученых.

Однако настоящим «отцом» отечественного искусственного интеллекта является профессор Поспелов Дмитрий Александрович, основатель научных школ ситуационного управления и семиотического моделирования, под «крышей» которых и развивался ИИ во времена, когда сам термин считался «ненаучным». Впервые официально и громко термин ИИ прозвучал в 1988 году, когда Д.А.Поспелов организовал Учредительный съезд ассоциации ИИ и была проведена первая всесоюзная конференция. Образовался научный совет ассоциации, куда вошли Попов Э.В., Осипов Г.С., Стефанюк В.Л., Хорошевский В.Ф., Городецкий В.И., Цейтин Г.С., Нариньяни А.А., Эрлих А.В., Гладун В.П. и другие известные специалисты.

В рамках данного комментария невозможно раскрыть и очертить всю широкую гамму результатов, полученных за 30 лет исследований. Заинтересованный читатель найдет все это в книгах, список которых прилагается. Мне хотелось бы только отметить следующее.

Проблемы, которыми мучаются герои Брайен и Бен - в ближайшие 100 лет не встанут перед учеными, поскольку сейчас даже близко никто не подошел к их подступам. Более того - в ту сторону никто и не идет. Моделирование эмоций на сегодняшний день звучит нелепо. Компьютер может лишь то, что закладывает в него человек, плюс быстрое действие процессора. Все. Психологи и физиологи слишком мало знают о человеке сейчас, чтобы говорить о моделировании процессов высшей нервной деятельности.

Привлечение идей «старика Фрейда» - еще один популистский шаг. Психология отличается сегодня тем, что ее многочисленные теории столь противоречивы и далеки от практических методик, что ни одну из них невозможно довести до минималь-

ной формализации, я уж не говорю о программной реализации. Имеющиеся же методики (тесты, опросники, игры и пр.), напротив, никак не подтверждены теорией, а используются на практике как некие инструменты, которые что-то меряют, но не ясно, что и как.

Автор перепрыгивает пропасть между первыми программами, которые основывались на идее решения задачи как поиска пути и перебора в лабиринте состояний (типично для шахматных программ), и современными системами, основанными на использовании баз знаний, тиражирующими опыт специалистов в конкретных узких прикладных областях (типа диагностика неисправности утюга), «забыв», что от идеи лабиринта отказались все, кроме создателей коммерческих игровых программ. Современные программы игры в шахматы, столь не «интеллектуальны», что это направление давно уже не обсуждается специалистами и исключено из программ всех научных конференций.

Столь пессимистические комментарии отнюдь не отрицают больших успехов сегодняшних интеллектуальных программных продуктов, ежегодный рынок которых превышает 400 млн. долларов США, и тех широких научных горизонтов, которые открываются перед современными исследователями. ИИ сегодня ставит больше вопросов, чем дает ответов. От исследователя требуется широкий кругозор, хороший междисциплинарный академический базис и ... здоровый скепсис. Однако настоящих ученых это не останавливает.

Если эта работа Марвина Минского в ком-то пробудила естественный интерес к проблематике искусственного интеллекта, то в Петербурге уже 7 лет действует городской научный семинар Петербургского отделения российской ассоциации искусственного интеллекта, который открыт для всех специалистов. Мы собираемся каждую последнюю среду месяца в 18-00 по адресу Гастелло, 12 (к.312 или 301). Справки по телефону: 293-32-68 у Гавриловой Татьяны Альбертовны (понедельник и четверг).

В Интернете наш адрес:

[http:// www.csa.ru/AI](http://www.csa.ru/AI)

Литература по ИИ, не требующая специальной подготовки:

1. Амамия М., Танака Ю. Архитектура ЭВМ и искусственный интеллект. - М.: Мир, 1993.
2. Будущее искусственного интеллекта. / Ред. составители Поспелов Д.А., Левитин К.Е./ - М.: Наука, 1991.
3. Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. - М.: Финансы и статистика, 1992.
4. Грановская Р.М., Березная И.Я. Интуиция и искусственный интеллект. - Л.: Изд. ЛГУ, 1991.
5. Лорьер Ж.-Л. Системы искусственного интеллекта. - М.: Мир, 1991.
6. Минский М. Фреймы для представления знаний. - М.: Энергия, 1979.
7. Мичи Д., Джонстон Р. Компьютер - творец. - М.: Мир, 1987.
8. Попов Э.В. Общение с ЭВМ на естественном языке. - М.: Наука, 1982.
9. Попов Э.В., Фоминых И.Н., Кисель А.В. Статические и динамические экспертные системы. - М.: Радио и связь, 1997.
10. Поспелов Д. А. Фантазия или наука? На пути к искусственному интеллекту. - М.: Наука, 1983.
11. Поспелов Д.А. Моделирование рассуждений. - М.: Радио и связь, 1989.
12. Справочник по искусственному интеллекту в 3-х т.т. /Под ред. Попова Э.В., Поспелова Д.А./ - М.: Радио и связь, 1990.
13. Уинстон П. Искусственный интеллект. - М.: Мир, 1990.
14. Уотермен Р. Руководство по экспертным системам. - М.: Мир, 1989.
15. Хейес-Рот Ф., Уотермен Д. и др. Построение экспертных систем. - М.: Мир, 1987.

*Воинов Александр Валерьевич,  
сотрудник Института прикладной  
астрономии РАН, кандидат  
технических наук.*

*Гаврилова Татьяна Альбертовна,  
доктор технических наук,  
профессор кафедры "Компьютерные  
интеллектуальные технологии в  
проектировании" СПбГУ,  
руководитель городского семинара  
по искусственному интеллекту.*

*Грановская Рада Михайловна,  
доктор психологических наук,  
профессор СПбГУ.*

*Иллюстрации к роману в тексте и  
на внешней стороне обложки  
Баранова Алексея Игоревича.*

НАШИ АВТОРЫ