

## ИСКУССТВО ИНТЕЛЛЕКТА

### УНИВЕРСАЛЬНАЯ МАШИНА МЫШЛЕНИЯ

Мы живем в удивительную эпоху, когда технологии, еще недавно казавшиеся уделом далекого фантастического будущего, вдруг оказываются чем-то обыденным. Особенно бурное развитие идет в области информационных технологий, в которой интеллектуальные системы выглядят наиболее фантастично. Так, еще 15 лет назад могло показаться сомнительным, что бытовые роботы, такие как роботы-пылесосы или газонокосилки, будут продаваться по всему миру каждый год миллионными тиражами, с методами машинного обучения будет незаметно для себя сталкиваться любой пользователь сети Интернет, каждый на своем телефоне сможет поставить систему дополненной реальности, использующую нетривиальные методы распознавания изображений. И даже автомобили, управляемые компьютером, если еще и не окажутся распространены повсеместно, то уже кое-где будут встречаться на дорогах.

### РОБОТЫ В НАШЕЙ ЖИЗНИ

Все это стало возможным, в том числе, благодаря развитию области искусственного интеллекта. Однако насколько менее фантастичной стала возможность создания самого искусственного интеллекта? Чтобы ответить на этот вопрос, нужно оглянуться назад на уже пройденный путь.

История всей техники – это история дополнения или замены частей человеческого тела искусственными приспособлениями: острый нож вместо мягких когтей, рычаг вместо слабых мускулов, колесо вместо медленных ног. В равной степени это касается и когнитивных функций – восприятия, памяти и самого мышления, для чего с давних времен использовались подзорные трубы, счеты, бумага с пером.

Однако в создании автономных механизмов, которые бы самостоятельно могли выполнять какие-то полезные (и, тем более, интеллектуальные) функции, успехи были гораздо скромнее. Конечно, механические музыканты, исполняющие без помощи человека какую-то мелодию, или часы, самостоятельно отсчитывающие время, существовали давно, но беспилотные транспортные средства или аналоги бытовых роботов до недавнего времени встречались лишь в сказках.

Что же изменилось? Вполне очевидно, что прогресс в области искусственного интеллекта, да и само существование этой области, обязано компьютерам. Более интересен, однако, вопрос, что революционного было в компьютерах? Ведь в XVIII–XIX веках шел бурный процесс автоматизации и без их использования. К примеру, изобретатели решали такие сложные задачи, как создание автоматов, выполняющих работу швей.

Подобные механизмы, однако, зачастую не были адаптивными и выполняли фиксированную последовательность действий.

Несколько более широким набором реакций обладали устройства, исходно призванные решать «интеллектуальные» задачи. Так, арифметическая машина, созданная Блезом Паскалем еще в XVII веке (а, возможно, известная людям и ранее), могла складывать или умножать любые числа, не закладываясь в нее на этапе конструирования. Тем не менее, ее возможности ограничивались заранее заданным набором операций. Понятно, что, чтобы поведение какой-то машины могло быть еще менее ограниченным, она должна быть способна выполнять новые для себя операции. Но как создать такую машину?

Ответ на этот вопрос удалось найти математикам, пытавшимся строго описать процесс решения математических задач, что привело к формальному определению понятия алгоритма, позволяющего в общем виде описать любые последовательности действий, выбор которых зависит от входных данных.

По сути, все автоматы прошлого оказались механическими реализациями тех или иных алгоритмов. Можно сказать, что изобретатели «программировали» на языке «шестеренок», не отделяя логическую структуру алгоритма от его механической реализации. Математика дала такое описание алгоритмов, по которому человек мог бы их сам выполнять «механически», не думая. Но может ли существовать такая машина, которая тоже сама будет выполнять любой алгоритм по его описанию? Как оказалось, может. И такую машину предложил Алан Тьюринг, доказав ее универсальность в 1947 году.

Ученые пытались придумать разные определения понятия алгоритма и универсальной машины, но все они оказывались эквивалентными (стоит отметить, что придуманы абстрактные модели так называемых сверхтьюринговых вычислений, но пока остается под вопросом возможность их физической реализации; по крайней мере, в отличие от классических алгоритмов, механически их человек воспроизвести принципиально не может). Эквивалентность здесь означает то, что одну универсальную машину можно описать как программу для другой машины и эмулировать работу первой

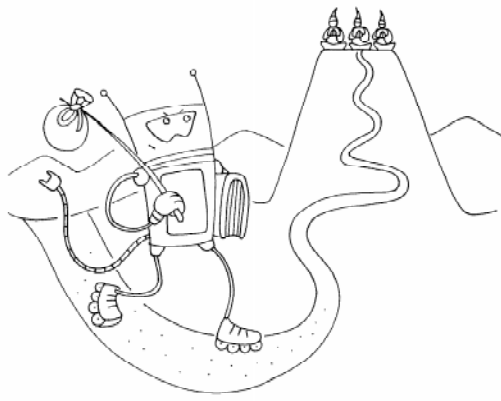
машины на второй. Стоит отметить, что, поскольку человек, снабженный ручкой и бумагой, может эмулировать действие универсальной машины Тьюринга, он сам является, по крайней мере, универсальной машиной. Это привело к выдвиганию тезиса Чёрча-Тьюринга, в соответствии с которым математическое и интуитивное понимание алгоритмов совпадают.

Для области искусственного интеллекта этот тезис можно интерпретировать так, что если мышление человека описывается алгоритмически (а никакого более широкого и притом физически реализуемого определения для программ поведения сейчас не известно), то оно может быть реализовано на универсальной машине. То есть человеческий интеллект не менее, но и не более чем универсальная машина.

Успех компьютеров, в частности (но далеко не только), в области искусственного интеллекта как раз и связан с тем, что они являются воплощением универсальных машин (а не просто быстрыми калькуляторами, как это нередко представляется). Для того чтобы компьютер выполнил новый алгоритм, не нужно внутри него переставлять местами шестеренки, как пришлось бы поступать изобретателю в XIX веке. И, главное, компьютер может выполнить любой мыслимый алгоритм (конечно, стоит иметь в виду то, что от абстрактной универсальной машины компьютер отличается ограни-



Рис. 1



*...магистральная ветвь развития области искусственного интеллекта шла путем повышения степени самостоятельности компьютера при решении задач.*

ченным объемом памяти и необходимостью работать в реальном времени).

Даже если принять то, что интеллект человека реализуется некоторой универсальной машиной, которая может эмулироваться на компьютере, остается вопрос, что это за машина, чем она отличается от других универсальных машин? Ведь универсальность в смысле способности выполнять любой алгоритм является, хоть и необходимым, но далеко не достаточным свойством интеллекта. В отличие от компьютера, человек не снабжается алгоритмами в явном виде, тогда как компьютер такой автономностью не обладает.

Можно сказать, что магистральная ветвь развития области искусственного интеллекта шла путем повышения степени самостоятельности компьютера при решении задач. На этом пути произошла смена ряда ключе-

вых аспектов интеллекта, находившихся в центре внимания исследователей.

### ЛАБИРИНТЫ ИНТЕЛЛЕКТА

На начальном этапе развития данной области господствовало представление о мышлении как поиске в пространстве решений. Ведь понятие алгоритма возникло в математике для описания обобщенной последовательности действий, позволяющей прийти от условия задачи к ее решению. Так, один алгоритм описывает то, как из двух чисел получить их произведение, а другой – как от значений коэффициентов квадратного уравнения перейти к его корням. Однако работа интеллекта заключается не в том, чтобы следовать известной последовательности действий, а в том, чтобы найти эту последовательность. К поиску последовательности действий сводится не только решение математических задач, но и игра в интеллектуальные игры. Да и психологические исследования, в том числе и в области поведения животных, это подтверждали. Так, Эдвард Торндайк еще в 1898 году писал, что решение задач, являющееся интеллектуальным актом, производится путем перебора различных манипуляций, действий. Мышление при этом уподобляется поиску пути в лабиринте, но лабиринте виртуальном, каждая развилка которого – это состояние решаемой задачи (например игровая ситуация или упрощаемое математическое выражение), а пути, выходящие из этой развилки соответствуют разным возможным действиям.

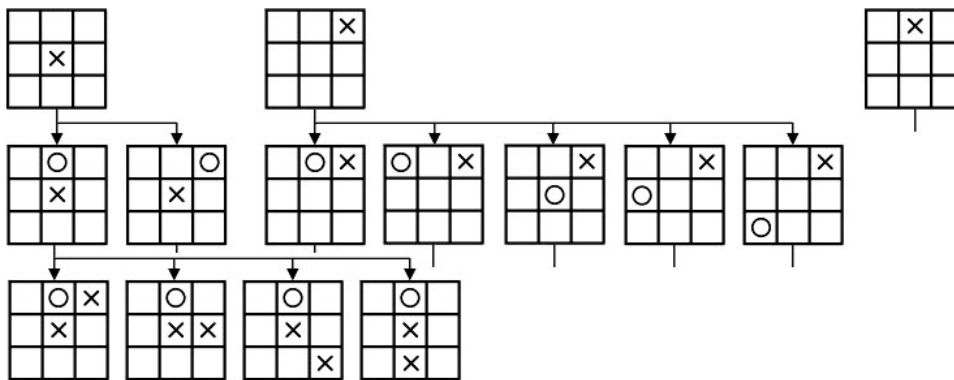


Рис. 2

## ДВИЖЕНИЕ ПО ДЕРЕВУ ИГРЫ

Но может ли для поиска решений совершенно разных задач существовать один и тот же алгоритм? Если имеется формальное описание задачи, то организовать поиск ее решения не так сложно: достаточно просто просматривать все допустимые цепочки операций и проверять, дают ли они решение. Однако при таком подходе сразу возникает проблема «комбинаторного взрыва»: число возможных цепочек очень быстро, экспоненциально, растет с ростом их длины.

Исследователи пытались решить эту проблему путем введения эвристик – приемов по сокращению перебора, которые позволяют отбросить неперспективные действия, но обычно не гарантируют достижения наилучшего результата. К примеру, простейшее правило сокращения перебора в шахматах, основывающееся на суммарной силе фигур, в первую очередь будет отбрасывать ходы, приводящие к потере ферзя, что обычно будет полезно, но также приведет к невозможности пожертвовать ферзем ради мата.

## ЗНАНИЕ – СИЛА

Как оказалось, при решении сложных задач лишь малая доля интеллектуальности (если судить по объему кода) приходится на общие методы поиска, а гораздо большая доля приходится на знания о предметной области. В итоге эвристические программы оказались способными решать только задачи узких классов, и под каждую задачу приходилось создавать свою программу. Компьютер удавалось наделять интеллектом, но только крайне узкоспециализированным. Для расширения возможностей компьютера нужно было, по крайней мере, избежать необходимости его программирования в явной форме под решение каждой задачи, для чего нужно было отделить предметные знания от механизмов рассуждений. Ведь мы можем согласиться с тем, что и человек не сразу может решать сложные задачи, а перед этим ему передаются соответствующие знания от других людей. Такую передачу знаний можно допустить и для искусственного интеллекта, но не через его прямое программирование.

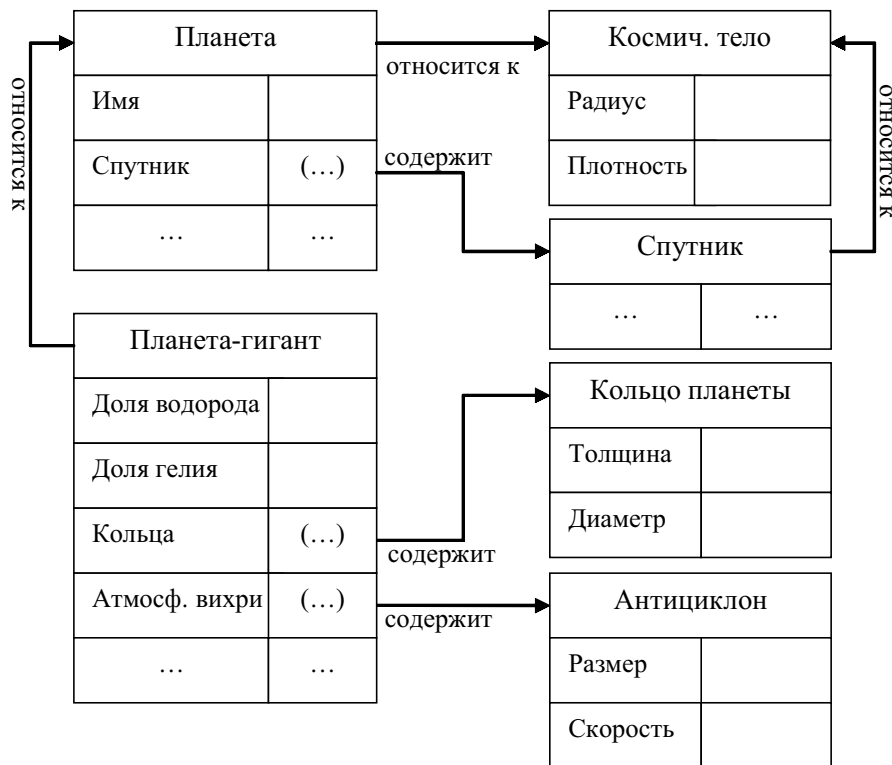
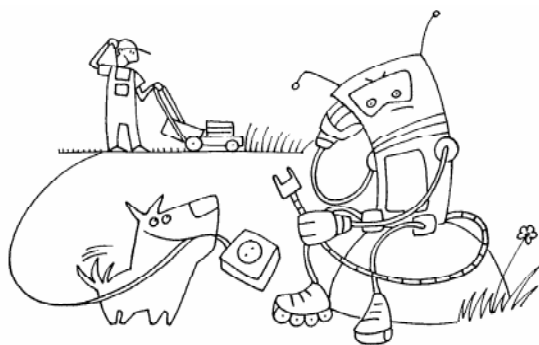


Рис. 3

## ПРЕДСТАВЛЕНИЕ ПРЕДМЕТНЫХ ЗНАНИЙ

Как результат, следующий этап развития данной области проходил под лозунгом «знание – сила». Здесь исследовались методы представления знаний, которые должны были позволить описывать различные предметные области так, чтобы интеллектуальные системы далее самостоятельно могли решать задачи из этих областей, будь то доказательство теорем, диагностика заболеваний или планирование действий робота. Идеи для представления знаний черпались из самых разных наук – психологии (ведь нужно было передавать знания компьютеру от человека), лингвистики (ведь наиболее удобным способом передачи знаний служит естественный язык), логики (ведь знания должны представляться так, чтобы над ними можно было производить автоматические рассуждения). Кроме того, правила поведения, общие понятия и связи между ними или сведения об устройстве конкретных объектов оказалось удобнее описывать в рамках разных представлений знаний. Вместо одного универсального представления знаний был разработан набор разнообразных инструментов по созданию экспертных систем – систем, основанных на знаниях и призванных решать задачи в узких предметных областях.

Наиболее трудоемким этапом при создании экспертных систем оказалось «извлечение» знаний из людей-экспертов с последую-



*...при решении сложных задач лишь малая доля интеллектуальности... приходится на общие методы поиска, а гораздо большая доля приходится на знания о предметной области.*

ющей их передачей компьютеру. Хотя этот процесс и не требовал явного программирования, он оказался нетривиальным, и выполняли его специальные люди – инженеры по знаниям. Ведь люди-эксперты приобретают знания нередко в неявной форме, на основе своего опыта, а не только от других людей, да еще и в форме строго сформулированных методических рекомендаций (которых зачастую как раз оказывается недостаточно).

Хотя экспертные системы оказались более самостоятельными (или универсальными), чем эвристические программы, их зависимость от человека оставалась огромной. Для ослабления этого ограничения компьютер должен был начать учиться на основе своего опыта или хотя бы извлекать знания из примеров решений, принимаемых людьми.

## УЧЕНЫЕ – СВЕТ

С 1980-х годов внимание исследователей в значительной мере смещается в область машинного обучения. Конечно, способность к обучению многими учеными и раньше считалась ключевой для интеллекта, но больший интерес представляли механизмы самого мышления как решения задач, пока попытки реализации этих механизмов не привели к упомянутым трудностям.

Стали развиваться методы машинного обучения, позволяющие обобщать примеры описаний конкретных ситуаций или принятых человеком решений с формированием общих понятий или правил для автоматического заполнения баз знаний. Соответствующие примеры, однако, готовились человеком и представлялись в символической форме. Такой источник обучающей информации ограничивал автономность интеллектуальных систем. Более того, рафинированная символическая информация не может передать всего богатства реального мира. К примеру, если какой-то интеллектуальной системе доступны только тексты на естественном языке, то понять их смысл, ни разу не «побывав» в реальном мире, не перемещаясь по нему, ничего не видя и не ощущая,



достаточно проблематично, а может, и вообще невозможно.

В свою очередь, методы машинного восприятия исследовались самостоятельно и во многом независимо от символьных методов, составлявших главный предмет изучения в традиционном искусственном интеллекте. Даже методы машинного обучения, активно изучавшиеся здесь еще с 1950-х годов на примере задач распознавания образов, оставались «субсимвольными». Эти методы, как нейросетевые, так и сугубо математические, зачастую сводились к аппроксимации отображения входных сигналов в выходные. Сама такая постановка является достаточно общей, так как в конечном итоге любой интеллектуальный агент учится преобразовывать входные данные в выходные команды. Однако сами методы машинного обучения до сих пор не являются универсальными.

Интересно, что общее теоретическое решение задачи машинного обучения как восстановления модели источника данных было предложено еще в 1960-е годы. Оно заключается в поиске наиболее короткого алгоритма, воспроизводящего данные наблюдений. Такой алгоритм и оказывается наиболее вероятной моделью источника данных. К сожалению, поиск наиболее короткого алгоритма – неразрешимая задача. Возможно, по этой причине о данном теоретическом результате практически нигде не упоминали, хотя он ясно показывает причину неуниверсальности известных практических методов

машинного обучения: все они используют тот или иной кусочек от целого множества алгоритмов, и, когда в данных содержатся непредусмотренные закономерности, они не могут быть точно представлены. Сейчас проблема универсального обучения привлекла заметно большее внимание, хотя до сих пор остается загадкой, как к нему можно приблизиться на практике. Однако и без универсальности методы машинного обучения играли и продолжают играть большую роль в искусственном интеллекте.

### ВОПЛОЩЕННЫЙ ИНТЕЛЛЕКТ

В 1980-е годы чисто символьные методы искусственного интеллекта подвергаются определенной критике. Как отмечалось выше, в традиционных системах, основанных на знаниях, смысл одних понятий выражается только через другие понятия, примерно как в словаре. Человек же способен понимать словесные определения новых понятий, поскольку на основе своего опыта в реальном мире уже понимает те понятия, на которых эти определения основываются. Такое «истинное» понимание оказывается недоступно системам, основанным на традиционных представлениях знаний. Кроме того, экспертные системы, даже дополненные функциями обучения, не обладают и многими другими способностями человеческого разума, такими как самосознание. Но, в то же время, эти системы проявляют опре-

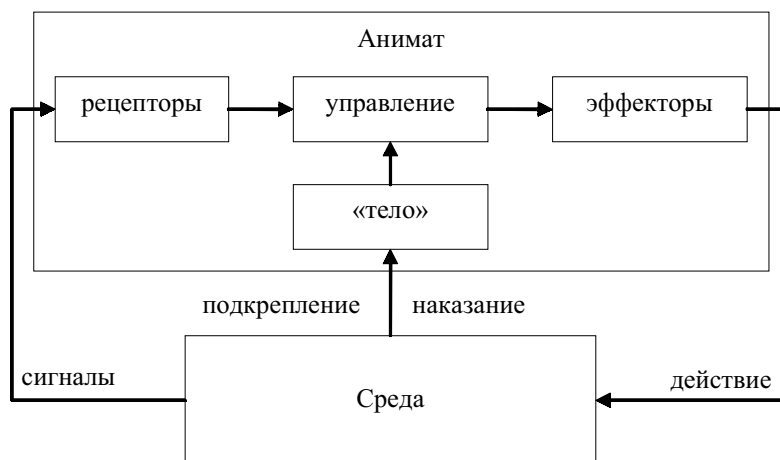
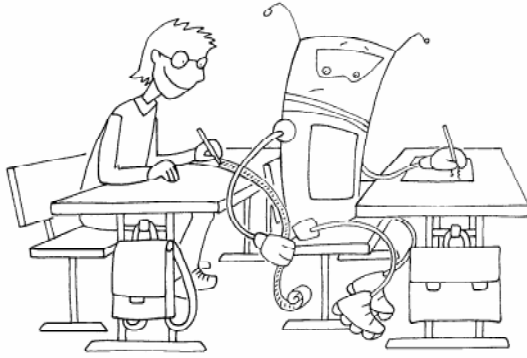


Рис. 4. Общая схема интеллектуального агента



*Стали развиваться методы машинного обучения, позволяющие обобщать примеры описаний конкретных ситуаций....*

*Соответствующие примеры, однако, готовились человеком...*

деленные интеллектуальные способности. Однако можно ли считать, что они обладают интеллектом?

В это время закрепляется разделение понятий сильного и слабого искусственного интеллекта. При этом под гипотетическим сильным искусственным интеллектом подразумевается настоящая разумная машина, тогда как к разряду слабых относят существующие интеллектуальные системы. В действительности, исследованием «настоящего» искусственного интеллекта уже толком никто серьезно не занимается. Это связано и с тем, что данная проблема оказалась гораздо сложнее, если вообще решаема, и ранние надежды на ее быстрое решение не оправдались, и с тем, что на практике оказываются востребованными системы, не наделенные самосознанием, а способные решать полезные задачи, и сделать такие системы удастся только весьма узкоспециализированными.

Тем не менее, стремление к ослаблению ограничений, характерных для чисто символьных систем, даже использующих обучение, но получающих информацию от человека, а не из среды, привело к значительному интересу к так называемому воплощенному интеллекту (к интеллектуальным автономным агентам, взаимодействующим с реальным миром), к обучающимся роботам, к проблеме автоматического формирования осмысленных понятий или связывания сим-

волов с сенсомоторным опытом, а также к устранению разрыва между символьными и субсимвольными методами.

## ОТ КОГНИТИВНЫХ АРХИТЕКТУР К ОБЩЕМУ ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ

Создание интеллектуальных агентов, способных воспринимать сенсорную информацию, обучаться, хранить знания о своем опыте, планировать и выполнять прочие когнитивные процедуры, причем выполнять эти процедуры нетривиальными методами, разработанными за десятилетия развития области искусственного интеллекта, привело к необходимости исследования способов объединения этих методов в единые системы, и с 2000-х годов популярной областью исследований становятся когнитивные архитектуры, как раз и призванные выполнить такое объединение.

Вместе с тем, начинает понемногу возвращаться интерес и к сильному искусственному интеллекту. Правда, теперь исследователи предпочитают противопоставлять не сильный и слабый, а общий и специализированный искусственные интеллекты, поскольку фактическим ограничением ранних интеллектуальных систем была именно их узкая специализация и зависимость от человека, тогда как такие субъективные качества, как самосознание или понимание, требуемые у сильного интеллекта, плохо определены, а их наличие у интеллектуального агента строго проверить невозможно.

Хотя проблематика общего искусственного интеллекта, первая конференция по которой прошла лишь в 2007 году, пока остается слишком радикальной, с точки зрения как академических, так и прикладных исследований, она является вполне логичным продолжением развития исследований в данной области, шедшей путем повышения универсальности систем искусственного интеллекта.

Однако будет ли этот этап, связанный с объединением всех предшествующих достижений искусственного интеллекта, последним? Вероятно, нет. Ведь объединяемые

слабые методы далеки от универсальности. Для методов поиска в пространстве решений остается в полной мере не разрешенной проблема комбинаторного взрыва. Представления знаний по-прежнему ограничены в своей выразительной силе. Практически применимые методы машинного обучения не способны выявить произвольные закономерности в данных или сформировать любые понятия, а теория универсального обучения остается слишком далека от практического использования. Хотя объединение слабых методов вряд ли позволит решить указанные проблемы, но совместное рассмотрение этих проблем может быть весьма полезным.

#### ОТ УНИВЕРСАЛЬНОГО ИСПОЛНИТЕЛЯ К УНИВЕРСАЛЬНОМУ ТВОРЦУ

В свое время математическая теория алгоритмов позволила создать универсальную машину, способную выполнять любые программы или эмулировать любую другую машину. Разрабатывая различные специализированные программы и передавая с их помощью часть своих интеллектуальных способностей компьютеру, человек создает себе автоматического помощника или заме-

ну в самых разных предметных областях. Чем интеллектуальнее программа, тем она автономнее, тем в более неопределенных условиях может работать и решать более сложные задачи. Самой же сложной задачей, обеспечивающей наибольшую автономность, является создание новых программ.

Эта задача не обязательно возникает как задача автоматизации труда программиста (хотя может быть рассмотрена и в такой постановке). Она также неразрывным образом связана с самим интеллектом – через формирование произвольных поведенческих программ, нахождение произвольных закономерностей в данных, представление произвольных знаний.

Задача построения машины, способной породить любой алгоритм (с желаемыми свойствами), неизмеримо сложнее задачи создания машины, способной выполнить любой алгоритм, но и влияние ее решения на нашу жизнь, видимо, будет еще больше. Однако пока остается совершенной загадкой, как эту задачу человеку сравнительно неплохо удается решать, в связи с чем сам интеллект, его использование и перенос в компьютер во многом пока остается искусством, как когда-то искусством было создание механических арифметических машин.

*Потапов Алексей Сергеевич,  
доктор технических наук,  
профессор кафедры компьютерной  
фотоники и видеоинформатики  
НИУ ИТМО.*



Наши авторы, 2013.  
Our authors, 2013.