

## **Основные подходы к созданию искусственного интеллекта**

*Лекция является кратким введением в учебный курс «Интеллектуальные системы» и раскрывает ключевые особенности информационного, бионического и эволюционного подходов к созданию искусственного интеллекта.*

### **Искусственный интеллект – как мы его себе представляем?**

Прежде чем говорить об особенностях подходов, нужно определиться с тем, *что* или даже *кто* скрывается за словами «искусственный интеллект» (сокращенно - «ИИ»). Учитывая высокую популярность «главного героя лекции», сегодня вряд ли найдется много студентов, которые не смогут привести примеры ИИ. Как правило, образ искусственного интеллекта ассоциируется с одним или несколькими персонажами художественных фильмов и фантастических романов, в которых они по замыслу автора играют роли:



- многофункционального и умного робота, помогающего решать бытовые задачи,
- бортового компьютера боевой машины или автопилота космического корабля,
- человекоподобного универсального солдата в виде робота, киборга или андроида,
- персонажа в компьютерной игре или другой сложной программы в виртуальном мире.

Несмотря на многообразие ролей у всех образов ИИ есть две противоречивые черты: с одной стороны, они обязательно чем-то отличаются от человека (внешним видом, характером мышления, набором чувств или т.д.), а с другой – стремятся быть максимально похожими на людей. Это противоречие авторы произведений ча-



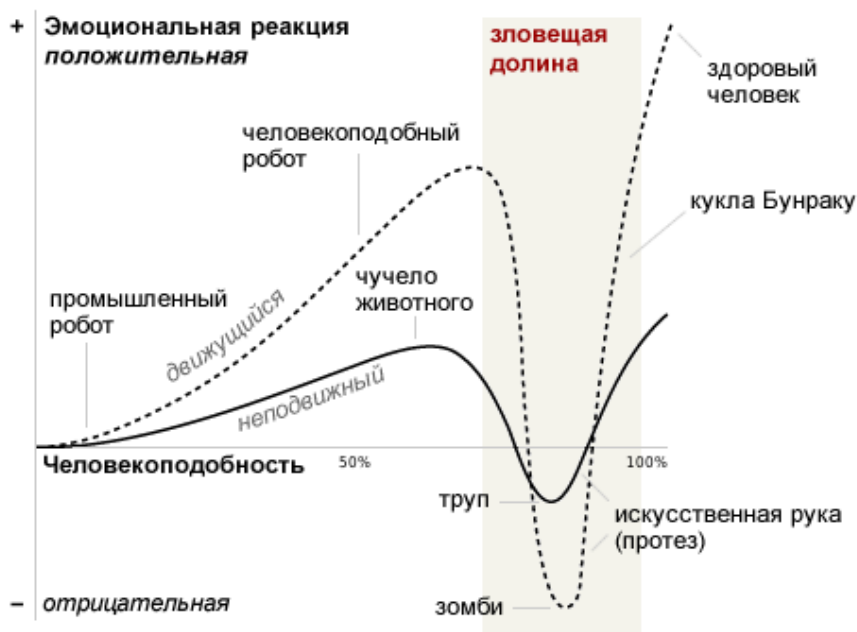
стенко используют как драматическую борьбу ИИ с самим с собой или окружающим его обществом.

Основной причиной и источником этой неоднозначности является стремление человека отделить себя от окружающего мира, выделить свои исключительные особенности, которые наиболее ярко проявляются в интеллекте и чувствах. Развитие интеллектуальных систем приводит к тому, что с каждым новым достижением машины «забирают» у человека его «привычную уникальность», а он стремится ее восполнить другими чертами внутреннего мира, все ближе отступая к иррациональной и часто необъяснимой грани своей сущности – чувствам, эмоциям, душе. Вынужденный характер такого отступления приводит к тому, что мы чаще воспринимаем ИИ негативно - со страхом или даже отвращением, несмотря на то, что он все ближе к нам по своему внешнему виду и внутреннему миру.



#### Интересный факт

В 1978 году японский ученый Масахиро Мори провел опрос, исследуя эмоциональную реакцию людей на внешний вид роботов. Поначалу результаты были предсказуемыми: чем больше робот похож на человека, тем симпатичнее он кажется — но лишь до определённого предела. Наиболее человекоподобные роботы неожиданно оказались неприятны людям из-за мелких несоответствий реальности, вызывающих чувство дискомфорта и страха. Неожиданный спад на графике «симпатии» был назван «Зловещей долиной» (англ. Uncanny Valley). Причина такого психологического феномена не выяснена до сих пор.



[\[http://en.wikipedia.org/wiki/Uncanny\\_valley\]](http://en.wikipedia.org/wiki/Uncanny_valley)

## Подходы к созданию искусственного интеллекта

В научной среде понятие «искусственного интеллекта» несколько отличается от распространенных образов научной фантастики и чаще всего трактуется не как конкретный объект/субъект, а как *междисциплинарная область передовых исследований информатики, робототехники, компьютерной и когнитивной науки*. Для обозначения результатов разработок в области ИИ используют понятия «интеллектуальная система», «интеллектуальный агент», «робот», «бот» и другие.

### Интересный факт

В русский язык понятие «искусственного интеллекта» пришло как результат не очень точного перевода словосочетания «*artificial intelligence*» (сокращенно - «AI»). Слово «*intelligence*» означает «способность рассуждать разумно» и не отождествляется с более емким и неоднозначным понятием интеллекта (англ. «*intellect*»). Под AI в англоязычной среде обычно подразумевается компьютерная программа или машина, которая обладает и демонстрирует интеллектуальные функции. Иногда, особенно в Англии, «*artificial intelligence*» используют как синоним понятия «*machine intelligence*».

Более подробно история возникновения понятия ИИ, словарные определения, ассоциации и толкования, а также основные направления и вехи развития области исследований рассматриваются в дополнительных лекциях, представленных на сайте учебного курса, в модуле «Начала искусственного интеллекта» - [http://it-claim.ru/Education/Course/AI/Lectons/Lectons\\_intro.htm](http://it-claim.ru/Education/Course/AI/Lectons/Lectons_intro.htm)

В рамках этой лекции рассмотрим основные подходы к созданию систем искусственного интеллекта, которые также будем называть *интеллектуальными системами*. Для наглядности сравнения различных подходов воспользуемся табличной классификацией из книги Питера Норвига и Стюарта Рассела<sup>1</sup>.

<b>Системы, которые думают как люди</b>	<b>Системы, которые думают рационально</b>
<b>Системы, которые действуют как люди</b>	<b>Системы, которые действуют рационально</b>

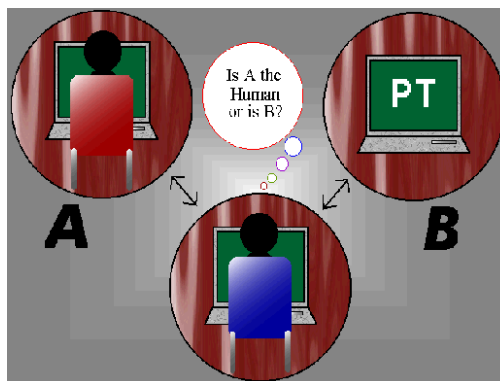
Формулировки в левом столбце соответствуют *бионическому направлению ИИ* и определяют интеллектуальные системы с точки зрения реализации человеческого мозга и организма. Определения в правом столбце раскрывают идеализированное (рациональное) понимание интеллектуальных функций и связывают с *информационным*

<sup>1</sup> Стюарт Рассел, Питер Норvig. Искусственный интеллект: современный подход (AIMA), 2-е издание, 2007 - 1408 стр.

направлением ИИ. Подходы, представленные в верхней строке, связаны с моделированием внутреннего мира человека - мыслительными процессами и рассуждениями. В нижней строке делается акцент на имитации интеллектуального поведения людей в окружающем мире.

### Действовать по-человечески: Тест Тьюринга

Тест, сформулированный Аланом Тьюрингом в 1950 г., был одной из первых попыток определить ИИ. Он предложил считать интеллектуальное поведение способностью достигать исполнения на человеческом уровне всех познавательных задач, достаточных, чтобы обмануть тестирующего. Суть теста заключается в том, что компьютер должен опрашиваться человеком через телетайп (чат). Если тестирующий не может определить компьютер или человек на другой стороне информационного канала, то компьютер можно считать интеллектуальным.



Для создания такого компьютера нужно над многим поработать - интеллектуальная система должна обладать следующими компонентами:

- *естественным языковым процессором* для успешного общения на некотором человеческом языке;
- *базой знаний* для хранения информации, полученной перед или в течение опроса;
- *пониманием*, чтобы использовать полученную информацию для ответов на вопросы и получения новых выводов;
- *машинным обучением*, чтобы приспосабливаться к новым обстоятельствам, обнаруживать и экстраполировать новые объекты (модели).

#### Интересный факт

В 1991 американский бизнесмен Лёбнер учредил специальную премию. Первой программе, которая пройдёт стандартный (текстовый) тест Тьюринга будет вручена серебряная медаль и вознаграждение в 25,000\$. Первой программе, которая пройдёт этот тест с использованием текстового, визуального и звукового подтверждения будет вручена золотая медаль и приз в 100,000\$. Кроме того, ежегодно лучшая программа получает приз в размере 2,000\$.

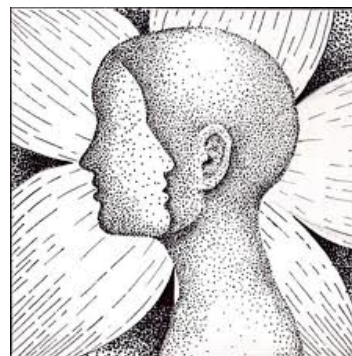
Тест Тьюринга умышленно исключает прямое физическое взаимодействие между тестирующим и компьютером, поскольку физическое моделирование человека не обязательно для интеллекта. Тем не менее, так называемый общий Тест Тьюринга

включает видео сигнал, чтобы тестирующий мог проверить способность ИИ воспринимать наглядную информацию и механизм передачи компьютеру физических объектов "через люк". Для того чтобы удовлетворить общему Тесту Тьюринга, компьютеру дополнительно нужно иметь:

- машинное зрение, чтобы воспринимать объекты;
- робототехнику, чтобы перемещать их.

### **Думать по-человечески: когнитивная модель**

Для создания интеллектуальной системы, которая может думать как человек, необходимо проникнуть в суть работы человеческого сознания. Есть несколько основных путей, чтобы сделать это: через *интроспекцию* (попытаться фиксировать собственные мысли в порядке их следования) или через проведение специализированного *психологического (когнитивного) эксперимента*. В последнее время когнитивные изыскания все чаще проводят вместе с исследователями мозга, пытаясь соединить достижения нейронауки и когнитивного моделирования.



Как только появится достаточно точная теория интеллекта и организации его познавательных функций, станет возможным ее реализация с помощью компьютерной программы. Информационные технологии могут служить эффективным инструментом для проверки гипотез или подтверждения их истинности. Кроме того, в настоящее время большой вклад в развитие когнитивных моделей вносят такие науки как *семиотика, лингвистика и психофизиология*.

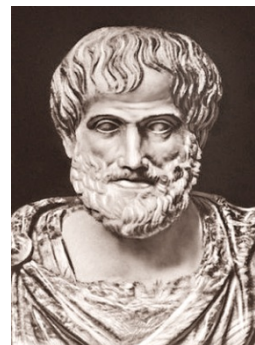
Значительные успехи в области компьютерных технологий привели к появлению междисциплинарной области когнитивной психологии, которая сводит вместе компьютерные модели ИИ и экспериментальные методы психологии, пытается создать точную и тестируемую теорию работы человеческого мозга. Последние тенденции и новейшие разработки в области когнитивных моделей и теории познания оказали влияние на выделение этой области в новую научную дисциплину, получившую название «*когнитивная наука*» или «*когнитология*».

#### **Интересный факт**

7 июля 2011 года Президент России подписал приказ, в котором определен перечень критических технологий Российской Федерации на ближайшие пять лет. В список вошли *Нано-, био-, информационные, когнитивные технологии (НБИК-технологии)*. Считается, что это направление работ является одним из наиболее перспективных и значимых в современной науке, а его достижения могут сильно изменить привычный нам мир.

### Думать рационально: законы мышления

Греческий философ Аристотель был одним из первых, попытавшимся определить «*правильное мышление*» как неопровержимые процессы рассуждения. Его знаменитые силлогизмы демонстрируют примеры структур, которые всегда делают правильные выводы при корректно заданных предпосылках. Например, "Сократ — человек; все люди смертны: следовательно Сократ смертен".



Есть два основных препятствия в этом направлении. Во-первых, неформальное знание не так легко представить в формальной форме, требуемой логической нотацией, особенно, когда знание определено менее чем на 100%. Для преодоления этого ограничения были разработаны специальные математические аппараты многозначных и модальных логик; теорий возможности, доверия, уверенности и др.

Во-вторых, есть большое различие между решением задачи "в принципе" и реализацией ее на практике. Проблемы даже с несколькими дюжинами фактов могут исчерпать все вычислительные ресурсы любого компьютера, если у него нет механизма, определяющего правильную последовательность шагов.

Хотя оба этих препятствия относятся к любой попытке сформировать вычислительные рассуждающие системы, они впервые появились в логике, где мощность представляющих и рассуждающих систем отчетливо выражена и хорошо понятна.

#### Интересный факт

Аллен Ньюэлл вместе со своим наставником Гербертом Саймоном разработал в 1956 году компьютерную программу Logical Theorist (Логик-теоретик), которая смогла автоматически доказать 38 законов из книги Рассела и Уайтхеда «Принципы математики» и даже предложить более элегантное доказательство одной из теорем. Это стало символическим достижением в развитии искусственного интеллекта.

### Действовать рационально: интеллектуальный агент

В рамках этого направления ИИ рассматривается создание и исследование интеллектуальных агентов - программ или роботов, которые воспринимают информацию и выполняют определенные действия. Действовать рационально предполагает оптимальное достижение заданной цели при имеющихся или заданных ограничениях.



Из «Законов мышления» Аристотеля следует, что единственный путь действовать рационально — это делать правильные логические выводы о том, что данное действие достигнет заданной цели, а затем действовать. Однако, такой подход не всегда возможен, т.к. существуют ситуации, где нет доказуемо правильной операции, которую нужно сделать.



Есть также способы действовать рационально, которые не включают логический вывод. Например, выдергивание руки из горячей печи: рефлекторное действие более успешно, чем медленное действие, выполненное после тщательного рассуждения.

#### **Интересные примеры робототехники**

- Андроиды Geminoid F и Geminoid HI-1, созданные выдающимся японским ученым Хироси Исигуро - <http://www.youtube.com/watch?v=9q4qwLknKag>
- Танцующая и поющая роботесса HRP-4C от фирмы Honda <http://www.youtube.com/watch?v=uP6Te5XbTF8>
- Японский робот ASIMO <http://www.youtube.com/watch?v=zul8ACjZI18>

#### **Цели создания искусственного интеллекта**

Попытка заставить машины действовать как можно более «разумно» привлекательна сама по себе, и как часто бывает, увлекаясь, многие исследователи, работающие в этой области, не делают серьезных попыток мотивировать свою деятельность. Однако, из приведенных выше толкований следует, что в исследованиях по ИИ все-таки можно выделить следующие основные цели.

Одна из них, ее часто называют *информационной* (эвристической, прагматической), состоит в создании программ для вычислительных машин, с помощью которых удалось бы автоматизировать такие виды человеческой деятельности, которые традиционно считаются интеллектуальными.



При этом, как будут построены эти программы, насколько близки или далеки будут те способы, которыми они достигают поставленной цели по сравнению с человеческими способами, абсолютно не важно. Имеет значение лишь конечный результат, его совпадение с тем, который получает человек при решении той же задачи.

Вторая цель — ее принято называть *бионической* — состоит в использовании программ искусственного интеллекта для объяснения умственной деятельности человека во

время решения тех или иных задач. В этом случае программы должны моделировать сам процесс получения результата человеком, помогать постигать эти процессы.

Для демонстрации успехов бионического направления ИИ, можно привести пример создания первой в мире бионической руки, которая помогает владельцу восполнить утрату настоящей кисти - [http://www.youtube.com/watch?v=OLkDQUOu\\_Y0](http://www.youtube.com/watch?v=OLkDQUOu_Y0)

Наряду с этими двумя направлениями существует и третье, которое называется *эволюционным*. Его представители считают, что интеллектуальные программы надо выращивать, как люди растят своих детей. Программы, по их мнению, только тогда станут по настоящему интеллектуальными, когда приобретут способность обучаться тому, чего они раньше абсолютно не умели делать. Целью при этом становится создание адаптирующихся, самообучающихся и самоорганизующихся систем.

#### **Цитата**

«Как бионический, так и эвристический подход к проблеме искусственного интеллекта носят в основном описательный характер, т. е. представляют собой попытки моделирования природы в том виде, в каком она нами наблюдается. В противоположность этому эволюционный подход носит, прежде всего, нормативный характер, так как является попыткой моделирования эволюционного процесса в том виде, в каком он мог бы происходить в природе, попыткой описания того, что должно было бы быть, вместо того, что есть. Ученый действует, чтобы знать, в то время как инженер знает, чтобы действовать. При эволюционном подходе исследователь в процессе умозрительного синтеза своей модели становится на позицию инженера...»

*Фогель Л., Оуэнс А., Уолш М.*

*Искусственный интеллект и эволюционное моделирование. - М.: Мир, 1969. -230 с.*

### **Укрупненные цели направлений в области ИИ**

<b>Бионическая</b>	<b>Изучение человека и его мышления, дублирование его биологических функций</b>
<b>Информационная</b>	<b>Автоматизация интеллектуальных функций человека, решение прикладных задач методами ИИ</b>
<b>Эволюционная</b>	<b>Понимание механизмов и условий формирования интеллектуальных функций в ходе эволюции и адаптации</b>



## **Возможность Искусственного Интеллекта**

Из того, что было представлено, можно сделать вывод, что за словами *искусственный интеллект* стоит некоторая важная *научная проблема*. Эта проблема выделяет область исследований и разработок искусственного интеллекта среди всех других работ в области вычислительной техники и программирования в самостоятельное научное направление. Проблему определяют два важных вопроса, остающихся открытыми:

- *Действительно ли методы, которые объединены в понятии искусственный интеллект, имитируют в значительной мере то, что мы понимаем под интеллектом?*
- *Существуют ли такие аспекты интеллекта человека, которые в принципе нельзя смоделировать на вычислительной машине?*

Оценки должны быть весьма осторожными, хотя бы с оглядкой на прогресс фантастических персонажей литературных произведений. Совсем недавно это были — гомункулусы, франкенштейны, а сейчас роботы, киборги, андроиды. Следует обратить внимание на все более детальное описание от одного литературного примера к другому основных функций современных роботов и киборгов. Такие «функциональные спецификации» все более походят на технические задания для инженерных разработок. При этом в основном специфицируются исключительно интеллектуальные функции этих существ путем вербального или графического описания.

В связи с этим *возможность искусственного интеллекта как объективная тенденция его развития* выглядит вполне определенно — *постепенная реализация в инженерных решениях одной за другой интеллектуальных функциональных способностей человека*.

Второй вопрос — это призыв к анализу искусственного интеллекта методом «от противного». В конкретных задачах, таких, как игра в шахматы, деятельность человека и машины поддается сравнению. Высказываются соображения, что существуют целые области деятельности человека, в которых машина не может даже и близко подойти к имитации человека. Наиболее яркий пример такой области деятельности человека — «новаторство», изобретательство, создание чего-то нового, ранее не существовавшего, заранее не predetermined.

Таким образом, *возможность искусственного интеллекта* следует рассматривать как конкретную (действительную) реализацию определенного набора функциональных возможностей интеллекта человека. При этом можно предположить некото-

рый «предельный», максимально полный набор, именуемый естественным интеллектом и к которому стремится научная и инженерная деятельность по реализации интеллекта искусственного.

Постановка вопроса о возможности реализации той или иной интеллектуальной функции равносильна постановке вопроса о ее познаваемости, поэтому *III возможен ровно настолько, насколько решен вопрос о познании интеллектуальных функций человека*. Все, что человек способен познать, можно в конечном итоге реализовать, а вопрос о механизме или форме реализации является вторичным.