

Искусственный интеллект общего типа:

состояние и перспективы

В статье дается анализ текущего состояния исследований и разработок в области искусственного интеллекта так называемого общего типа, сравнимого с человеческим. Рассмотрены текущие проблемы и анализируются подходы к их решению.



Основными направлениями в научных исследованиях по теме «искусственный интеллект» являются:

- Artificial Intelligence (AI, искусственный интеллект), представляющее совокупность из приблизительно 25–30 научных дисциплин, относящихся к областям автоматизации решения отдельных задач, ранее доступных только человеку, как то: распознавание образов; автоматический перевод; формальная логика (игры, интеллектуальное взаимодействие с человеком на естественном языке) и др.;

- Artificial General Intelligence (AGI, искусственный интеллект общего типа) — программно-аппаратное решение, которое автономно может решать интеллектуальные задачи в совокупном перечне и с качеством, присущим человеку.

Краткий обзор состояния отрасли AI

Сделать полновесный обзор состояния отрасли ИИ на текущий момент — задача архисложная. И тем не менее, к 2019 году программы, относящиеся к направлению «искусственный интеллект», доказали доминирование над человеком в настольных знаковых играх, причем не только в априорно детерминированных (шахматы, го и т.п.), но и в изначально неопределенных, типа покера. Постепенно зона доминирования интеллектуальных программ расширяется на профессиональную деятельность. Так, программа IBM Watson продемонстрировала 90% достоверности консультаций по юридическим вопросам на фоне 60% со стороны экспертов-адвокатов. Сама по себе точность информационных систем никого удивить не может, но в данном случае корректно обрабатываются именно неформализованные запросы. Конечно, в этом примере следует также учесть прецедентность американского права, что, безусловно, сильно осложняет работу любого адвоката в США.

Беспилотные автомобили перешли от этапа исследований в фазу внедрения, хотя до их массового использования еще очень далеко. По мнению автора, это еще лет 25.

Также впечатляющие успехи ИИ-отрасль демонстрирует в области медицины и биологии. Эти успехи связаны, прежде всего, с обработкой больших данных (big data) и анализом изображений. Так при автоматическом анализе кардиограмм на искусственных нейросетях достигнут уровень достоверности обнаружения отклонений и идентификации основных заболеваний сердца в 80% и более случаев. И это при затратах времени на анализ — 10 секунд! При диагностике, например, некоторых видов рака по рентгеновским снимкам человек уже не может конкурировать с теми же, соответствующим образом обученными, нейросетями, прежде всего, учитывая фактор времени.



Автомобиль на автопилоте — уже реальность

В области гражданской безопасности распознавание лиц уверенно занимает все смежные ниши и работает практически в реальном времени на библиотеках любых размеров. Хотя это относится только к фронтальному представлению и заданному разрешению/масштабу. Идентификация лиц при ракурсной съемке остается нерешенной проблемой.

Автоматические переводчики типа «текст–текст» дают практически на 90% приемлемый подстрочник. И качество перевода постоянно улучшается.

Автоматические трансляторы «речь–текст» уже вполне в «уличной практике» и используются, прежде всего, в варианте «английский–китайский».

В области информационной робототехники «помощники-секретари» уже вполне адекватны, по крайней мере в статистически массовых запросах, и уже вышли на рынок через смартфоны.

Автономные робототехнические комплексы «топчутся» в клиниках — как «помощники», а в общественных местах — как «робокопы», но они уже уверенно «зарабатывают деньги» на складах и постепенно внедряются в доставку — на базе авиационных дронов.

В военной сфере продвижение, с одной стороны, не может быть столь публичным, с другой — можно предположить, что оно не столь впечатляюще. Причина этого кроется в том, что никто сегодня не отважится признать, что решение о летальной атаке принимается автоматически. Вернее, команда на пуск ракеты или сброс бомбы, конечно, формируется системой, но на сегодня — только при наличии «разрешения на летальное применение», полученного от оператора-человека. Кроме того, военная сфера требует совершенно иного уровня испытаний и сертификации, на что еще просто не было достаточно времени.

Из военных применений за последнее время можно выделить создание практическими всеми ведущими странами приложений, связанных с массовым применением воздушных дронов на основе т.н. swam intelligence («интеллекта роя» или «коллективного интеллекта»). Подобные приложения



Рой дронов — яркая иллюстрация применения систем коллективного искусственного интеллекта в военной сфере

на базе подводных автономных устройств созданы, похоже, только в США.

Автономные боевые дроны реально пока не летают... Они падают... Летают только при удаленном управлении: со спутника, с параллельного истребителя, с земли в режиме прямой видимости. До полностью автономных полетов еще далеко. Вопрос тут намного глубже, чем это может показаться. Дело в том, что управление самолетами создавалось с учетом фактора пилота-человека! Это позволяет резко понизить порядок систем дифференциальных уравнений, подлежащих решению в режиме «он-лайн». Так, например, поведение военно-транспортного самолета на этапе посадки в целом описывается системой из 24 дифференциальных уравнений! С учетом наложенных ограничений, за которые потом отвечает пилот (ну сейчас еще и автопилот, но все-таки с учетом наличия и пилота), порядок системы удастся свести до 4-х уравнений. Но когда сейчас занялись автономной авиацией, необходимо вернуться к базовой динамике полета самолетов. А это непростая задача.

Если говорить о беспилотных автомобилях, то перспективная задача их внедрения не только не усложнится — она сильно упростится! Действительно, зачем автономному устройству светофор и, соответственно, необходимость распознавать его сигнал? Достаточно просто излучать в ближней зоне перекрестка три частоты и не «заморачиваться» с техническим зрением... Ситуация с беспилотными автомобилями очень напоминает время, когда в конце 19-го и начале 20-го века они только появились на дорогах. Там шли люди, ехали лошадиные повозки и между ними лавировали «самодвижущиеся устройства» с ограничением скорости в 6 км/час! Автомобили не стали учить распознавать лошадей! Просто реорганизовали пространство и минимально технически его оснастили. Вероятно, это произойдет и в случае беспилотных систем.

Большая часть работ в области AI выполняется на основе искусственных нейросетей. Поразительно, как много может дать в решении конкретных практических задач понятие «линейного пространства» и «линейных операторов», которые являются математической основой данного направления. Можно только представить, чего мы достигнем, когда разовьется нелинейный анализ!



Зачем автономному автомобилю светофор?

Однако применение нейросетей требует очень затратного обучения. При этом уже обученная нейросеть практически не может быть «переучена»! Но говоря о возможностях развития, нас, как минимум, интересуют самообучающиеся и переобучающиеся автоматы. А вот до этого пока далеко.

Следует отметить, что привыкая к демонстрируемым успехам в области ИИ, мы постепенно отказываемся от признания указанных сфер таковыми, которые можно было бы отождествить с искусственным интеллектом. Достижения эти становятся привычными и обыденными (сейчас никого не удивить программами, которые играют в шахматы на уровне чемпиона мира и даже лучше). А нам все время кажется, что ИИ — это нечто, которое придет к нам в будущем и которое пока что недостижимо ни в плане практической реализации, ни даже в части осмысления.

Более того, в этом будущем мы будем постоянно сталкиваться скорее с тем, что именуется **«искусственным интеллектом общего типа»**, и где-то в чем-то оно будет даже обладать некими задатками разума. Собственно, с одной стороны, речь вовсе не идет о создании искусственного разума, хотя это достаточно странно. Но с другой стороны, зачем нам такой AGI, если он по уровню не будет дотягивать до разума хотя бы среднего человека. Ведь стоимость такой системы намного превышает затраты на обучение обычного человека. Разработки имеют смысл, если предполагается создание интеллекта, во много раз превосходящего человеческий. Потому что интеллект человеческого уровня уже достигнут, будем считать, естественным путем.

А вот получить инструмент, намного превосходящий возможности человеческого ума — это действительно сверхзадача, решение которой выведет цивилизацию совершенно на другой уровень. Но одновременно с этим такое решение будет невероятно опасным инструментом, используя который можно будет успешно и быстро реализовывать различные негуманные цели. То государство, которое первым создаст высокоуровневый AGI, гипотетически будет доминировать во всех мыслимых сферах — от военной и освоения космоса до медицины, науки управления обществом. Это будет эволюционный прорыв с последствиями, которые трудно спрогнозировать и даже не хотелось бы предсказывать.

А посему немудрено, что работы по созданию AGI, начиная с определенного момента, будут строжайше засекречены. Хотя то, что известно двоим, все равно скоро станет известно третьему. И никакие договоренности здесь не помогут — ставки слишком велики.

Работы в направлении AGI

Что же сегодня можно сказать о разработке **искусственного интеллекта общего типа**?

Постановка вопроса немедленно требует конструктивных определений таких понятий, как «интеллект», «сознание» (consciousness), «мышление» (thinking), «познание» (cognition). В приложении к человеку эти его свойства исторически изучались в рамках научного направления Cognitive Science — «наука о познании». Оно возникло задолго до

изобретения компьютеров и было практически всецело посвящено изучению интеллекта человека. Но с момента начала компьютеризации задачу создания «машинного разума» стали обозначать термином «искусственный интеллект». В настоящее время можно отметить, что на фоне скромных успехов в разработке автономных интеллектуальных машин значение «науки о познании» возрастает, и работы по созданию искусственного интеллекта общего типа все более связываются с ожидаемым прогрессом именно в данном научном направлении.

Какое же наследство нам принесла «наука о познании» и что в ней сейчас происходит? Успехи в развитии направления оптимального управления в 50–60-е годы прошлого века, которые совпали с появлением компьютеров, немедленно привели к возникновению исследовательского сегмента «Вычислительная теория мышления» (Computational Theory of Mind — СТМ). В основу данного подхода была положена технически очень продуктивная концепция теории управления, которая состоит в том, что для сложных систем выходной сигнал должен обязательно зависеть от состояния системы, а не только от его входного воздействия.

При этом данный подход называется также **семантическим**, так как предполагается, что компьютер оперирует **символами** как на уровне операций, так и на уровне выходных представлений. Отдельный и очень непростой вопрос при этом — каким образом компьютер сможет сформировать символическое представление сложных понятий окружающего мира? Либо придется предположить, что в рамках данного подхода некоторый **понятийный набор** будет предоставлен машине априори — т.е. он будет наследоваться при ее создании, а остальные понятия будут уже конструироваться самостоятельно.

И в этом нет ничего необычного. Ведь если взять, к примеру, математику, то понятие **множества** также относится к не определяемым понятиям. И вообще, в основе математики лежат утверждения, которые базируются лишь на интуиции, которая скорее и определяет развитие этой науки. Все новые направления начинались с конструктивных построений, а только потом они аксиоматизировались и формализовались. Попытка Давида Гильберта построить строгую и непротиворечивую символическую математику, в которой бы все базировалось только на строгом следовании правилам операций с символами, провалилась.

Однако, по-видимому, ограничения, которые возникнут у искусственных интеллектуальных систем в смысле **познаваемости мира**, не будут являться для них критическим препятствием.

Человек, как единичное существо, при всем его индивидуальном интеллекте не создал бы современную науку без: 1) социального взаимодействия посредством языка, 2) фиксации знаний посредством языка и письменности, 3) системы образования, 4) возможности освобождения некоторых ресурсов общества в виде его специально отобранных представителей («ученых») от необходимости все свое время тратить на добывание пищи и обеспечение собственной жизнедеятельности, а в современном обществе — полного



ИИ общего типа (AGI) требует четких конструктивных определений

освобождения ученых от необходимости материальной продуктивной деятельности.

Когнитивная наука

В этой части изложения автор в основном будет придерживаться структуры направления, представленной в работе «Michael R.W. Dawson Mind, Body, World: Foundations of Cognitive Science, AU Press, Athabasca University, 2013».

В переводе Cognitive science упоминается как «наука о познании», «когнитивистика», «когнитология», «когнитивные науки», «наука о мышлении», т.е. в русском языке термин еще нельзя назвать устоявшимся.

Cognitive science (CS) как научное направление возникла в 50-е годы прошлого века в виде попытки преодолеть фрагментацию в исследовании процессов познания в существующих различных дисциплинах.

В связи с появлением в это время компьютеров направление CS отнесли к области **обработки информации**.

В результате довольно сложного взаимодействия перечисленных дисциплин само направление CS стало фрагментированным.

По классификации, общепринятой среди исследователей в области искусственного интеллекта, Artificial Intelligence обозначает совокупность направлений, связанных с моделированием отдельных интеллектуальных функций. Отсюда можно заключить, что к CS относится именно направление исследований по созданию искусственного интеллекта общего типа.

В настоящее время в CS выделяют следующие основные направления:

- классическое;
- коннекционистское;
- «телесное».

Прежде всего данные направления расходятся в трактовке понятия «обработка информации» (**information processing**).



Человек не создал бы науку без освобождения некоторых ресурсов в виде его специально отобранных представителей («ученых») от необходимости все свое время тратить на добывание пищи

Классическая CS интерпретирует этот термин как «манипуляцию символами на основе предписанных правил» — в таком же смысле, как это положено в основу цифровых компьютеров. Однако впоследствии ряд исследователей выразили сомнения, что искусственный интеллект общего вида удастся создать просто в результате написания хороших программ. Можно сказать, что отчасти критики классической CS оказались правы, учитывая хотя бы тот факт, что до сегодняшнего дня прогресс в создании AGI очень скромный.

Однако в области решения корректно поставленных (в математическом смысле) задач классический подход продемонстрировал потрясающие успехи, прежде всего в моделировании игр «по правилам» (шахматы, го) и «не очень по правилам» — таких, как покер, где устраивать чемпионаты с компьютерами человеку уже нет смысла. В корректно поставленных задачах удается точно описать и пространство состояний, и пространство целей, что позволяет проектировать эффективные алгоритмы управления. К неудачам классического подхода относятся задачи из области распознавания образов и естественной речи, как в понимании смысла, так и в поддержании диалога.

Коннекционистская CS предлагает модель, полностью отличную от архитектуры цифрового компьютера — она базируется на структурах, в основе которых лежит нейронная



Интеллект присущ всему живому, хоть в какой-то степени обладающему инстинктом самосохранения, который выражается в адапционных возможностях организма к изменениям окружающей среды

организация мозга. При отсутствии достоверных знаний о законах обработки информации в мозгу исследователи базируются в основном на модели нейрона, в котором организована весовая обработка сигналов, поступающих по нервным волокнам (аналог синапсов). Удивительным достижением является то, что нейросети удается обучать с достижением вероятности правильного распознавания, превышающей 80% даже на изображениях, которые трудно классифицируются даже человеком, например: рентгеновские снимки опухолей или сигналы кардиограмм. Недостатки коннекционистского подхода заключаются в том, что:

- не удается создать теорию обратной связи; это приводит к тому, что в настоящее время используются только «прямые» искусственные нейросети;
- обучение на «размеченных» выборках является дорогостоящим процессом;
- переучивание нейросети хоть и допускается, но реально этим никто не занимается, так как нет никакой гарантии, что новая задача будет оптимально решаться в той же архитектуре связей;
- организация последовательного алгоритмического процесса на нейросети пока не осмыслена.

Сложившаяся ситуация приводит к тому, что корректно поставленные задачи в настоящее время решаются на основе классического направления CS, а некорректные, которые сводятся к распознаванию образов и допускают возможность обучения по тренировочным выборкам, решаются с помощью технологии искусственных нейросетей.

Третье направление CS — т.н. **«телесный» подход** (embodied cognition approach), возникший в конце 80-х — начале 90-х годов прошедшего века. Аргументом телесного подхода является то, что инструментом познания у человека очевидно является не только мозг, но и все тело. Взаимодействие тела с окружающим миром дает мозгу значительно больший объем информации, чем тот, который он мог бы получить, используя только зрение и слух. Апологеты телесного подхода считают, что цикл «ощущение–мысль–действие» (sense–think–act), присущий классическому направлению CS, должен быть заменен непрерывной обработкой процесса «ощущение–действие» (sense–act processing).

Ощущение и действие при этом рассматриваются не в последовательности, а в едином процессе. Мозг при этом, как считается, является скорее управляющим агентом цикла «ощущение–действие», чем его планировщиком. Телесный подход в CS наследует у кибернетики идею прямого взаимодействия с окружающей средой и адаптацию к ее изменениям, что и является процессом познания.

Таким образом, в противовес вычислительному подходу была выдвинута теоретическая концепция, базирующаяся на следующих тезисах:

1. Познание телесно, или «отелесненно»; то, что познается и как познается, зависит от строения тела и его конкретных функциональных особенностей, способностей восприятия и движения в пространстве. Устроено по-разному — познается по-разному.

2. Познание ситуационно. Познающее тело погружено в более широкое — внешнее природное и, в случае человека, социокультурное окружение, оказывающее на него свое влияние.

3. Познание осуществляется в действии, через действия животной особи; через действия формируются и когнитивные способности, как видовые, так и индивидуальные; когнитивная активность в мире создает и саму окружающую по отношению к познающему существу среду — в смысле отбора, вырезания им из мира именно только того, что соответствует его телесным потребностям, когнитивным способностям и установкам.

4. Познавательные системы есть динамические и самоорганизующиеся системы. В этом функционирование познавательных систем принципиально сходно, единолично функционированию познаваемых природных систем, то есть объектов окружающего мира.

Также в рамках телесного подхода предлагается «расширенное понимание мышления, которое означает, что этот процесс осуществляется только в среде пребывания, а не ограничено только мозгом.

Автор во многом придерживается «телесного» подхода и развивает его. Однако антагонизм между тремя приведенными направлениями, как мне представляется, несколько драматизирован — скорее всего, они будут взаимодействовать, а никак не вытеснять друг друга.

В последние годы автор последовательно представил видение тезауруса искусственного интеллекта, включая гипотезу иерархичности интеллекта и определений некоторых понятий: «Искусственного интеллекта индивидуального и коллективного типов», «Мысли», «Эмоции» и ряд других. Отличием представленных определений является то, что они могут непосредственно служить конструктивным целям и программироваться уже сегодня, чем и планируется заняться в ближайшие годы.

В основу определения интеллекта индивидуального положен инстинкт самосохранения индивида — с точки зрения автора эти понятия по сути тождественны. В основе интеллекта коллективного — инстинкт самосохранения популяции; в наиболее широком представлении — вида. Ранее в «СиБ» были опубликованы статьи автора: **«Искусственный интеллект: как и зачем?» («СиБ», №2, 2015)**, **«Искусственный интеллект как иерархическая структура» («СиБ», №3, 2016)** и ряд других. Научные статьи автора представлены на его аккаунте в исследовательской социальной сети ResearchGate, ссылка на который приведена в конце статьи.

Базовым понятием в авторском определении интеллекта индивидуального является понятие «граница». Дело в том, что понятие «тело» требует точного математического определения, которого нет. Поэтому автор для развития телесного подхода — его математизации — использовал известное математическое понятие **границы**.

О «проблемах языкознания»

Вернемся к классическому подходу. Он во многом опирается на понятие языка. Предполагается, что это средство формулировки мыслей. Речь, понятно, идет не о языке, на котором этот интеллект программировали, хотя этот вопрос тоже не из последних... А речь в данном случае о том, должен ли этот интеллект обладать языком? Если да, то каким и для чего?

Какую роль язык выполняет в человеческом обществе? Это формулировка знаний, их хранение и передача. Язык здесь выступает как средство коммуникации. Все эти функции являются социальными! Полностью автономный субъект, вообще-то, может функционировать и без языка — например, те же животные. Хотя они не совсем без языка, но он у них присутствует в довольно упрощенной форме и с сокращенными функциями. Но вопрос в том, собираемся ли мы создавать социальную среду интеллектуальных роботов? Хотим ли мы, чтобы они накапливали и передавали знания? Хотим ли мы, чтобы они воспроизводили себе подобных и воспитывали их?

Язык робота — его «первичный язык» — это цифры. Они позволяют осуществлять автоматическую классификацию и просто нумеровать классы — причем без ограничения словаря! Для робота количество слов не обязано ограничиваться 16 тысячами, как для Шекспира, или 12–14 тысячами, как для Пушкина, — словарь робота принципиально может иметь любую размерность!

Интерфейс с человеком

Другой вопрос — взаимодействие человека и интеллектуальных роботов. Во многих случаях человеку невозможно будет понять, что именно ему объяснили без специального ПО интерпретации — программы «переводчика» с языка робота на естественный язык человека. Основная причина — система понятий робота будет другой, чем у человека! И «думать», соответственно, он будет иначе.

Другое дело, что сейчас выделено отдельное направление т.н. «коллаборационных роботов». Это интеллектуальные системы, которые должны взаимодействовать с человеком. Такие устройства, соответственно, должны понимать естественную речь и общаться на ней с человеком.

Что по этому поводу можно сказать? Прежде всего, язык робота — «интеллектуального агента» — не может совпадать с языком человека! Язык робота должен быть другой!



С созданием «Искусственного интеллекта общего вида» все, конечно, не столь печально и беспросветно, как может показаться на первый взгляд: автономные автомобили ездят по дорогам; беспилотники летают все лучше и лучше; «помощники-ассистенты» помогают и ассистируют; искусственные нейросети анализируют и диагностируют, распознают и обнаруживают, сортируют и стоят на страже! Отрасль создания приложений искусственного интеллекта



К вопросу интерфейса между человеком и компьютером

Почему? Да просто потому, что его сенсорная система, в общем случае, иная: радиолокационная, лазерная, инфракрасная, химическая. В общем-то, она может быть реализована в любом из освоенных частотных диапазонов.

Стакан наполовину полон... или наполовину пуст?

Не все так плохо, как кажется. И не все так хорошо, как хотелось бы.

Что происходит с направлением «искусственный интеллект» в социальном смысле? Для обычного человека научные статьи по проблемам искусственного интеллекта уже давно интуитивно не понимаемы — нужно знать не просто математику, но и конкретные особенности рассматриваемого AI-направления, следить за новинками научных публикаций — «быть в теме». На вопрос, который иногда задают приятели: объясни, пожалуйста, «что такое искусственный интеллект», «когда он появится» и «как будет устроен» автору уже давно сказать нечего... Это как попытаться коротко и обиходно ответить на вопросы: «расскажи коротко о теории чисел», или «многомерных пространствах», или о «топологии»... Создание искусственного интеллекта общего типа, как и развитие всех направлений, объединяемых понятием AI, — это критически работа математиков! Не программистов, не психологов, не философов, не юристов, не экономистов, хотя все эти профессии участвуют в процессе и вносят свою лепту в данное научное направление. Но руководить работой будут математики. Возможно, с физиками..., имея в виду квантовые компьютеры. Математизация направления произошла давно, и это, конечно, обособило его от возможности «бытовых дискуссий»: тема стала сложной и скучной — она стала профессиональной!

(направление AI) уже работает вовсю! Достижения в сфере «искусственного интеллекта общего типа» (направление AGI) не столь заметны, но автор может заверить читателя, что «бикфордов шнур» тлеет последние сантиметры...

Сергей КОРНЕЕВ,
специалист по искусственному интеллекту,
системам связи и управления
https://www.researchgate.net/profile/Sergey_Korneev2
<https://ua.linkedin.com/in/sergeykorneyev>