

# Дмитрий Поспелов

## Десять «горячих точек» в исследованиях по искусственному интеллекту

Как самостоятельное научное направление искусственный интеллект (ИИ) существует чуть более четверти века. За это время отношение общества к специалистам, занимающимся подобными исследованиями, претерпело эволюцию от скепсиса к уважению. В передовых странах работы в области интеллектуальных систем поддерживаются на всех уровнях общества. Бытует устойчивое мнение, что именно эти исследования будут определять характер того информационного общества, которое уже приходит на смену индустриальной цивилизации, достигшей своей высшей точки расцвета в XX-м веке.

За прошедшие годы становления ИИ как особой научной дисциплины сформировались ее концептуальные модели, накопились специфические, принадлежащие только ей методы и приемы, устоялись некоторые фундаментальные парадигмы. Искусственный интеллект стал вполне уважаемой наукой, ничуть не менее почетной и нужной, чем физика или биология.

У специалистов старшего поколения, стоявших у колыбели зарождения новой сферы исследований, складывается убеждение, что период бурного развития кончился и теперь наступает эра вполне академических, спокойных и целенаправленных исследований, рассчитанных на длительный период.

Поэтому было бы весьма любопытно попытаться увидеть те основные направления исследований в ИИ, те «горячие точки», в которых будут сосредоточены основные усилия специалистов в конце уходящего века и начале нового тысячелетия. Анализ состояний текущих исследований позволяет выдвинуть предположение о наличии десяти таких «горячих точек». Именно они и будут ниже вынесены на обсуждение.

### 1. Переход к аргументации

В начальный период развития ИИ идея применения механизмов логического вывода в аксиоматических (или квазиаксиоматических, использующих в качестве аксиом определенные законы данной предметной области) системах занимала доминирующее положение. Предполагалось, что все или почти все задачи, претендующие на интеллектуальность, можно решать путем построения некоторого вывода. Такая парадигма породила многочисленные работы в области автоматического доказательства теорем, разработки языков представления знаний логического типа, в частности, хорошо известного языка Пролог. Значительные усилия были затрачены на создание методов вывода в исчислении предикатов, которое различным образом модифицировалось, чтобы адаптировать его для нужд искусственного интеллекта.

Классический подход в ИИ, реализующийся под явным давлением логических моделей в представлении знаний, породил экспертные системы, основанные на продукционных правилах, теорию реляционных баз данных, теорию решателей и планировщиков. Несомненным преимуществом, связанным с увлечением логическим выводом, было привлечение в сферу исследований области ИИ логиков, принесших в эту молодую науку свои представления о строгости и точности постановок задач и формулировок результатов.

Но уже к середине 70-х годов постепенно выясняется, что классических логических моделей и схем вывода явно не хватает для того, чтобы строить достаточно богатые и практически значимые интеллектуальные системы. Искусственный интеллект явно вырос из «логических штанишек». Принципы, опирающиеся на классическое понимание формальной системы дедуктивного вывода, стали слишком узкими для решения задач ИИ. Возникло нечто вроде кризиса в физике, ярко проявившегося в начале XX-го века. В чем же состояла основная проблема?

Логический подход в его классической форме требовал для каждой предметной области, для которой применялись методы ИИ, наличия полного перечня исходных положений, которые можно было бы считать аксиомами этой предметной области. Их существование (сюда естественно включаются и априорно задаваемые правила вывода) обеспечивало замкнутость используемых моделей, позволяло ставить и решать круг проблем, связанных с полнотой, результативностью и непротиворечивостью используемых моделей и процедур.

Однако различные приложения, к которым стремился искусственный интеллект, оправдывая свою практическую значимость, в подавляющем большинстве случаев не давали возможностей построения аксиоматических систем. Знания о предметных областях, как правило, были неполными, неточными и лишь правдоподобными, что приводило к эффектам немонотонности процессов получения результатов, возникновению фальсификаторов ранее полученных утверждений, быстрому снижению достоверности утверждений, получаемых в результате последовательного (даже при так называемых параллельных модификациях) процесса логического вывода.

Так возникла проблема замены формальной системы с присущими ей процедурами дедуктивного вывода иной, столь же мощной моделью, где отражались бы основные особенности поиска решения в плохо определенных предметных областях, которые описываются как открытые системы с обновляемыми знаниями об их строении и функционировании.

С конца 70-х годов XX-го века старая парадигма, опирающаяся на идею строгого логического вывода, начинает постепенно сменяться новой парадигмой, провозглашающей, что основной операцией при поиске решения должна быть правдоподобная аргументация. Работа с аргументами «за» и «против», снабженных соответствующими весами, приводит к аддитивным процедурам с этими весами (в противовес мультипликативным процедурам вычисления обобщенных весов при правдоподобном выводе). Это обстоятельство оказалось решающим для перехода к аргументации в интеллектуальных системах.

Однако, в отличие от завершенной структуры логического вывода, до сих пор не существует столь же стройной, научно разработанной теории правдоподобной аргументации. Эффекты, связанные с появлением парадоксов при немонотонных рассуждениях, показывают, что переход к более богатой по сравнению с моделью логического вывода модели правдоподобной аргументации неизбежно приводит к большому количеству новых проблем, связанных с обоснованием подобной модели и изучением ее особенностей. Предполагается, что в ближайшие десятилетия усилия многих специалистов сосредоточатся именно в этой области исследований.

## **2. Проблема оправдания**

При заполнении памяти интеллектуальных систем знаниями, полученными от экспертов, хорошо знающих данную предметную область и способы решения возникающих в ней задач, инженеры по знаниям столкнулись с одной весьма любопытной особенностью. При объединении баз знаний, сформированных различными экспертами, каждая из которых в отдельности довольно успешно решала тестовые задачи, возникала противоречивая база знаний. Если это были базы, в которых знания фиксировались в виде продукционных правил, то возникали правила с одинаковыми левыми частями и противоречащими друг другу правыми частями, а если в качестве модели представления знаний использовались фреймы, то отмечалось несовпадение фреймов-прототипов или возникали конкурирующие значения в слотах.

Как скоро стало очевидным, явление это объяснилось тем, что эксперты погружали свои профессиональные знания в некоторые концептуальные модели предметных областей, которые могли не совпадать у разных экспертов. Если, например, речь шла об извлечении знаний из области онкологических заболеваний, то в качестве таких различных моделей

могли выступать вирусная или генетическая модель порождения злокачественных опухолей. Если же речь шла о добыче нефти и поиске наиболее перспективных мест для ее разведки, то альтернативными концептуальными моделями оказывались модели органического и неорганического происхождения нефти. Приверженцы той или иной концептуальной модели сознательно (а иногда подсознательно) отбирали только те знания, которые согласовывались с принятой концептуальной моделью, отвергая те знания, которые ей противоречили (или, преуменьшая их вес правдоподобия). Другими словами, использовался своеобразный механизм «психологической защиты» от знаний, разрушающих принятую концептуальную модель, которая оправдывалась принятой системой знаний.

Ясно, что механизмы оправдания, связанные с искажением весов правдоподобия, накопленных в данной предметной области знаний (включая и отторжение некоторых знаний), нужно изучать и учитывать при заполнении баз знаний и их объединении. Но как это делать? Вопрос остается открытым и сейчас. В ближайшее десятилетие на него надо найти конструктивный ответ, иначе интеллектуальные системы в своем развитии не сделают следующего важного шага.

### **3. Порождение объяснений.**

Различие в механизмах поиска решений у человека, специалиста по решению определенного класса задач и у интеллектуальной системы приводит к появлению эффекта непонимания. Видя окончательный результат деятельности интеллектуальной системы, специалист не может оценить степень его достоверности (даже если система одновременно с решением выдает вычисленную ею такую оценку). Процесс «верить – не верить» не может привести к какому-либо разумному результату, если нет какой-то дополнительной информации. Эта информация, начиная с первого поколения экспертных систем, выдается специалисту по его требованию в виде объяснения.

Объяснение (более точно было бы говорить об обосновании, но термин «объяснение» прочно прижился в искусственном интеллекте) может быть различным. Наиболее распространены как-объяснения и почему-объяснения. При как-объяснении система выдает пользователю информацию о процедуре получения решения, например, выдает ему всю трассу движения по дереву вывода. При почему-объяснении система включает в текст объяснения те основания, которые были использованы ею в процессе поиска решения. Другие типы объяснений (что-объяснения, зачем-объяснения и т.п.) пока в интеллектуальных системах практически не используются.

Проблема объяснения связана с решением задачи о том, как его построить. В начальный период развития экспертных систем объяснение порождалось жесткими процедурами, связанными с наличием априорно заданного сценария объяснения, заполняемого конкретными сведениями в процессе поиска решения. Такой подход в какой-то мере годится для как-объяснений и несколько хуже подходит для почему-объяснений. Но он совершенно непригоден при необходимости порождения объяснений иных типов.

Ясно, что проблема генерации объяснений связана с наличием в памяти интеллектуальной системы концептуальных моделей предметных областей и энциклопедических знаний об особенностях этих областей. Для порождения объяснений нужны базы знаний энциклопедического типа.

Энциклопедические базы знаний отличны по своей структуре и назначению от проблемно-ориентированных баз знаний, применяемых в современных экспертных и других интеллектуальных системах. Организация знаний в энциклопедических базах, механизмы порождения текстов по запросам к ним требуют решения ряда проблем, относящихся к форме представления знаний в таких базах и принципам построения системы отношений на хранимых в них знаниях.

С середины 80-х годов начали выполняться первые проекты, связанные с созданием энциклопедических баз знаний и развитием их теории. Без накопления новых результатов в

этом направлении процесс совершенствования интеллектуальных систем и снятия барьера непонимания между ними и пользователями-людьми будет вряд ли возможным.

#### **4. Поиск релевантных знаний.**

Базы знаний в интеллектуальных системах используются в самых различных процедурах, связанных с вводом знаний, преобразованием и использованием их для поиска решений. Все эти процедуры требуют наличия механизмов, которые позволяли бы по тексту запроса к базе знаний найти в базе именно те знания, которые понадобятся для получения ответа.

В отличие от баз данных, для которых механизмы поиска ответов на запросы достаточно хорошо изучены (хотя и тут еще имеется немало нерешенных задач, являющихся для специалистов «твердыми орешками»), соответствующие проблемы для баз знаний пока еще далеки от каких-либо эффективных решений. Известно, что основным механизмом реализации процедур поиска в базах знаний служит поиск по образцу, являющийся своеобразным видом ассоциативного поиска.

Процедуры поиска по образцу имеют высокую сложность с экспоненциальной оценкой. Но с этим приходится мириться, так как пока не найдено какой-либо иной универсальной процедуры, позволяющей отыскивать нужные знания. Но основной проблемой остается поиск ответа на вопрос: «Как сформировать образец по тексту поступившего запроса?».

Сложность этой проблемы можно проиллюстрировать на следующем (весьма простом) примере. Пусть в систему введен текст: «Петя залез на стул, открыл дверцу заветного шкафа, достал конфеты и быстро, торопясь, съел несколько штук. ...Вечером мать, доставая конфеты для чая, посмотрела на Петю и сказала: „Странно. Конфет стало меньше“. Петя густо покраснел».

Предположим, что интеллектуальная система должна ответить на вопрос: «Почему Петя покраснел?» На основании какой информации она сможет это сделать? Как мы сами отвечаем на подобные вопросы?

Проблема поиска релевантной некоторому тексту информации пока остается нерешенной. В этой области пока больше вопросов, чем ответов. В ближайшие годы новые поколения ученых должны внести в решение этой проблемы свою лепту. Возможно, что появление энциклопедических баз знаний облегчит эту задачу.

#### **5. Понимание текстов**

Что значит понять текст? Ответ на этот вопрос, приемлемый в равной степени для лингвиста, психолога или философа, отсутствует. Феномен понимания во многом еще остается загадочным. В интеллектуальных системах имеется способ уточнить интерпретацию этого термина. Можно, например, считать, что система понимает введенный в нее текст, если с точки зрения некоторого человека (или группы экспертов) она правильно отвечает на все вопросы, связанные с информацией, отраженной в тексте.

Такая интерпретация связана с классификацией типов вопросов, которые, в свою очередь, определяют уровни понимания текста. На первом уровне все вопросы прямо связаны с предложенным текстом и ответы на них в явном виде содержатся в этом тексте. Если использовать вышеприведенный текст про Петю, съевшего конфеты, который был приведен выше, то вопросами первого уровня могли бы быть, например, следующие: «Куда залез Петя? или «Что сказала мать?». Вопросы второго уровня требуют специального логически получаемого расширения текста. Ибо ответы на них в явном виде в тексте отсутствуют. Пример такого вопроса: «Почему конфет стало меньше?». Ответ на такой вопрос опирается на некоторую схему рассуждения. Третий уровень понимания связан с расширением текста за счет привлечения релевантных знаний. Пример такого вопроса был приведен при обсуждении соответствующей проблемы. Но возможны и более высокие

уровни понимания. Например, уровни, связанные с тем, что текст понимается не буквально, а метафорически. Особым случаем понимания текста является вычленение из него прагматической составляющей, некоторого руководства к действию. Такая интерпретация процесса понимания тесно связана с задачей планирования поведения на основе текстового описания условий и целей.

## 6. Синтез текстов.

С проблемами поиска релевантного знания и понимания текстов весьма тесно соприкасается еще одна проблема, привлекающая сейчас внимание специалистов, работающих в ИИ. Она связана с разгадкой механизмов, лежащих в основе процедур порождения текстов на заданную тему. Без ответов на возникающие тут вопросы нельзя организовать полноценный обмен информацией между людьми и системами искусственного интеллекта. Ибо механизм генерации целенаправленного текста вместе с механизмом анализа и понимания текстов образует основу процесса коммуникации – главного процесса в организации человеческого поведения и реализации всех видов его деятельности.

Как и процесс понимания, процесс синтеза текстов имеет многоуровневую структуру. После зарождения мотивов его генерации и осознания целей, которые предполагается достигнуть в акте общения, наступает этап порождения когнитивной структуры текста. Этот этап реализуется на уровне внутренних представлений системы о мире, хранящихся в базе знаний. Знания, релевантные целям, которые направляют процесс «строительства» текста, отбираются некоторым планировщиком на знаниях и собираются во внутреннюю структуру текста. После этого на уровне лингвистической компоновки текста другой планировщик превращает эту структуру в линейный текст на естественном языке. Этот текст еще лишен того, что в психолингвистике называют читабельностью. Он еще слишком связан с машинными представлениями. Куски плохо пригнаны друг к другу, отсутствует гладкость переходов и ясность изложения целей. Эти недостатки «глубинного текста» исправляются на третьем уровне генерации, который реализуется стилистическим планировщиком.

Описанная процедура генерации ставит перед ее создателями ряд проблем, не решенных к настоящему времени. Например, неясно, какие принципы лежат в основе построения когнитивных структур текстов. В пользу того, что такие принципы существуют, убедительно свидетельствуют, например, эксперименты по генерации текстов волшебных сказок или музыкальных произведений (также текстов, но использующих специальный язык для общения с пользователями). В первом случае когнитивная структура определяется набором глубинных функций В.Я. Проппа, задающих достаточно жесткий сценарий будущего текста. Во втором случае имеется конечная система правил, делающих процедуру композиции в нужной мере формальной.

Но остаются, по крайней мере, две важные проблемы, решение которых пока не найдено: а) как цели связаны с когнитивной структурой и б) как описываются когнитивные структуры тех типов текстов, которые нужны, например, в процессе естественно-языкового диалога. Большое внимание к проблемам теории речевых актов (нового направления в лингвистике) со стороны специалистов по ИИ подогревается надеждами найти здесь ответы на поставленные вопросы.

Неменьшей проблемой является переход от нелинейной структуры текста к ее линейному представлению. Этот переход тесно связан с исследованиями по гипертекстам. Определенный бум, возникший в этой области, как раз и связан с осознанием того факта, что линейный по форме текст, как правило, является внешним кодом нелинейной структуры, на которую он «натянут». Гипертекстовые технологии призваны не только обеспечить возможность работы с нелинейным представлением текстов, но и должны как-то решать задачи его линеаризации и перехода от линейного представления к гипертекстовому.

Этот комплекс взаимосвязанных задач сейчас настолько активно изучается, что есть немалые основания считать, что в ближайшие годы проблемы синтеза текстов найдут свое

разрешение.

## 7. Когнитивная графика

Исторически сложилось так, что системы технического зрения и машинной графики всегда находились где-то на окраине области ИИ. Как и модели распознавания образов, методы, используемые для решения возникающих здесь задач, по своей сути были мало чем похожи на те, которые традиционно использовали специалисты по искусственному интеллекту. Для классических «систем, основанных на знаниях», как часто называются экспертные и другие интеллектуальные системы, уровень сенсорных и перцептивных процессов, играющих фундаментальную роль при зрительном восприятии или восприятии речи, оказался слишком «мелким». В их базах знаний был реализован куда более «крупный» уровень ментальных представлений. И пока специалисты по использованию зрительной и акустической информации в интеллектуальных системах занимались «нижними» уровнями восприятия и генерации, остальные специалисты, работающие в области ИИ, не находили с ними общего языка.

Настоящее общение между ними началось в 80-е годы, когда стали появляться первые исследования в области ментальной интерпретации перцептивных образов (анализ трехмерных сцен) и в области анимации зрительных картин, связанных с ментальными представлениями. Установление связи между текстами, описывающими сцены, и соответствующими изображениями потребовало наличия в базах знаний специальных представлений для зрительных образов и процедур соотнесения их с традиционными формами представления знаний.

Графическая информация стала трактоваться с позиций знаний, содержащихся в ней. Если до этого ее функция сводилась к иллюстрации тех или иных знаний и решений, то теперь она стала включаться равноправным образом в те когнитивные процессы, которые моделируются в базах знаний и на основе их содержимого. Термин «когнитивная графика» отражает этот принципиальный переход от иллюстрирующих изображений к видеообразам, способствующим решению задач и активно используемых для этого.

Когнитивная функция изображений использовалась в науке и до появления компьютеров. Образные представления, связанные с понятиями граф, дерево, сеть и т.п. помогли доказать немало новых теорем, круги Эйлера позволили визуализировать абстрактное отношение силлогистики Аристотеля, диаграммы Венна сделали наглядными процедуры анализа функций алгебры логики.

Систематическое использование когнитивной графики в компьютерах в составе человеко-машинных систем сулит многое. Даже весьма робкие попытки в этом направлении, известные как мультимедиа-технологии, привлекающие сейчас пристальное внимание специалистов (особенно тех, кто занят созданием интеллектуальных обучающих систем), показывают перспективность подобных исследований.

Пока же область компьютеризации правополушарных функций мозга человека остается почти terra incognita. Здесь начаты лишь первые большие проекты, направленные на создание систем, опирающихся на когнитивную графику. На наш взгляд, в ближайшие годы следует ожидать качественного прорыва в этой области ИИ.

## 8. Многоагентные системы

Тема с таким названием возникла на конференциях, посвященных проблемам ИИ, где-то в первой половине 80-х годов. Причин для появления такой проблематики было несколько. Прежде всего, стало ясно, что эффективная реализация ряда важных для интеллектуальных систем процедур требует параллельной и асинхронной их организации. Подобные процессы интегрируют в себе активности отдельных центров, решающих свои локальные задачи. Но эти локальные задачи и пути их решения должны быть согласованы в

границах некоторых глобальных целей.

Примерами процедур такого рода могут быть процедуры согласования мнений различных экспертов по поиску решения сложной многоцелевой задачи, согласование локальных локомоций при синтезе интегрального движения (например, движение робота, снабженного зрением и манипуляторами) или процедура коллективного взаимодействия интеллектуальных систем при решении в автономном режиме некоторой общей задачи.

Появление специальных архитектур, призванных поддерживать такую организацию процессов (например, параллельные вычислительные системы, в которых используется принцип «доски объявлений»), еще более усилило интерес к многоагентным моделям. Наконец, уверенность в том, что в нервных тканях живых организмов реализуется асинхронный и параллельный режим поиска решения, также оказала свое влияние на исследования в области многоагентных систем.

Нужно отметить, что идеология моделей такого рода во многом опирается на методы и результаты, полученные ранее вне сферы интересов собственно искусственного интеллекта. Еще в конце 50-х годов появились первые работы в области клеточных автоматов и моделей коллективного поведения автоматов. Эти работы заложили основу для появления многоагентных систем. Новое, что внесли в эти исследования специалисты по интеллектуальным системам, – это повышение «уровня интеллекта» агентов. Они стали способны использовать свои локальные знания для достижения своих целей. И задачи согласования, организации их целесообразного взаимодействия трансформировались на верхнем уровне в задачи согласования целей и знаний, т.е. стали напрямую соотноситься с проблематикой искусственного интеллекта.

Возникающие тут проблемы тесно связаны с проблемами динамических баз знаний, с необходимостью оценки конфликтных целей, противоречий в знаниях. Они также предполагают использование упоминавшихся выше процедур оправдания в системах имеющихся знаний и концептуальных моделей.

Сторонники этого нового системного движения надеются, что в начале следующего века будет создано новое научное направление – теория асинхронных конфликтующих процессов или что-то подобное с другим названием, которое еще не появилось.

## **9. Сетевые модели.**

Интеллектуальные системы, основанные на правилах (продукциях), принесли не только радость решения ряда важных задач, но и породили сомнения в том, что именно они призваны остаться основными моделями представления знаний в интеллектуальных системах. Многочисленные дискуссии 80-х годов, проводившиеся специалистами в области ИИ по этому поводу, привели к укреплению сетевой парадигмы, несколько отодвинутой в сторону триумфальным выходом на сцену продукционных моделей. И хотя исследования в области семантических сетей, каузальных сетей и сетей другого типа продолжались, они были малочисленными и не слишком продуктивными.

Но к концу 80-х годов сетевые модели стали развиваться более быстрыми темпами. Этот процесс совпал с пробуждением интереса к давно забытым нейронным архитектурам, появлением транспьютерных систем и нейрокомпьютеров, а также с возвращением к работам, опирающимся на эволюционные модели и эволюционное программирование. Возник определенный бум, который был даже окрещен неodarвинизмом.

Если к концу первого этапа развития сетевых моделей (в основном в виде нейронных многослойных систем типа персептронов) наступило разочарование в их возможностях и простоте их аппаратной реализации, то в 80-х годах эти сомнения были отброшены. Комплекс исследований в этой области так возрос, что произошло практическое отпочкование специалистов, работающих в области сетевых моделей, от основного ядра тех, кто причисляет себя к искусственному интеллекту. У «сетевиков» появились свои журналы, они стали проводить свои симпозиумы и конференции и формировать свою терминологию.

Этот разрыв нарастает, что по-видимому, приведет к возникновению двух наук, связанных с построением интеллектуальных систем. Одна из них будет по-прежнему опираться на уровень ментальных (информационных) представлений, а другая – на уровень структурной организации (по типу нервных тканей), порождающей нужные решения. Во всяком случае в 90-е годы вряд ли можно ожидать спад интереса к сетевым моделям и многочисленным нерешенным проблемам, связанным с их построением и функционированием.

## **10. Метазнания.**

Метазнания или знания о знаниях – неперенный атрибут познавательных процессов. В искусственных системах они в том или ином виде присутствовали всегда (например, в виде схем баз данных в базах данных или в виде стратегий управления в производственных системах).

Но только с полным осознанием глобальной цели искусственного интеллекта, которую можно сформулировать, как создание метасистемы, способной порождать все необходимые конкретные программы деятельности, стало ясно, что уровень метазнаний сам по себе представляет немалый интерес для изучения. Метазнания тесно связаны с теми основными для человека процедурами, которые позволяют ему учиться новым видам деятельности. Именно поэтому интерес к метазнаниям тесно связан с глубоким вниманием к процессу обучения, которое характерно для начала 90-х годов.

Интеллектуальные обучающие системы, использующие метазнания для организации учебного процесса, ориентированного на конкретного обучаемого, стали первым объектом, в котором метазнания «овеществились», приобрели все необходимые качества для конкретного изучения. В 90-х годах мы, наверняка, станем свидетелями первых впечатляющих результатов в этой области.

Заканчивая эту статью, хочу подчеркнуть, что выбор описанных тут десяти «горячих точек» исследований в области искусственного интеллекта, конечно, субъективен. Другие специалисты могли бы назвать и другие важные направления в развитии интеллектуальных систем. Но я тешу себя надеждой, что пересечение их с упомянутыми в статье направлениями было бы значительным.