

АКАДЕМИЯ НАУК СССР  
ИНСТИТУТ ИСТОРИИ ЕСТЕСТВОЗНАНИЯ И ТЕХНИКИ

На правах рукописи

Герович Вячеслав Александрович

ДИНАМИКА ИССЛЕДОВАТЕЛЬСКИХ ПРОГРАММ  
В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ДИССЕРТАЦИЯ

на соискание ученой степени кандидата философских наук

специальность 09.00.08. – философские вопросы естествознания и техники

Научный руководитель: доктор философских наук А.А. Печенкин

Москва – 1991

## ОГЛАВЛЕНИЕ

Введение	3
Глава 1. Методология науковедческого анализа области искусственного интеллекта	9
1.1. Область искусственного интеллекта как объект науковедческого анализа	9
1.1.1. Постановка проблемы	9
1.1.2. Искусственный интеллект: инженерия или наука?	10
1.1.3. Гносеологическая характеристика области	14
1.2. Анализ применимости различных схем развития науки к истории искусственного интеллекта	17
1.2.1. Общая характеристика логико-методологических схем К. Поппера, Т. Куна и И. Лакатоса	17
1.2.2. Парадигмы в истории искусственного интеллекта	19
1.2.3. Научно-исследовательские программы в истории искусственного интеллекта	22
1.3. Характеристика адекватной модели развития области	24
1.3.1. Сходство со схемой И. Лакатоса	24
1.3.2. Отличительные особенности модели	26
Глава 2. Историческая динамика области искусственного интеллекта	31
2.1. Истоки и зарождение исследовательских программ	31
2.1.1. Кибернетический контекст	31
2.1.2. Зарождение исследовательских программ	33
2.2. 1960-е годы: смена задач	38
2.2.1. Кризис "традиционного" искусственного интеллекта	38
2.2.2. Искусственный интеллект в естественных условиях	41
2.3. 1970-е годы: конкуренция исследовательских программ	43
2.3.1. Проект DARPA	43
2.3.2. Раскол научного сообщества	45
2.4. 1980-е годы: наука или коммерция?	48
2.4.1. Инженерный бум	48
2.4.2. "Булевская" и "небулевская" исследовательские программы	50
Глава 3. Роль рефлексии ученых в развитии исследований по искусственному интеллекту	55
3.1. Особенности рефлексии научного сообщества	55
3.1.1. Методологические проблемы анализа научной рефлексии	55
3.1.2. Моделирование научного познания	56
3.1.3. Модели рефлексии	58
3.2. Виды деятельности и типы рефлексии в области искусственного интеллекта	59
3.3. О роли вопроса "Может ли машина мыслить?"	63
3.3.1. Мироззренческий смысл	63
3.3.2. Практический смысл	65
3.3.3. Гносеологический смысл	67
Заключение	69
Список литературы	73

## ВВЕДЕНИЕ

Актуальность темы определяется значимостью исследований по искусственному интеллекту, находящихся на переднем крае информатики и комплекса когнитивных наук. Проблема искусственного интеллекта является поистине уникальной по широте и разнообразию привлекаемых для ее решения научных знаний и инженерных приемов, психологических моделей и философских идей.

Являясь продуктом сложных процессов интеграции и дифференциации в современной науке, исследования по искусственному интеллекту оказывают непосредственное воздействие на смежные области (психологию, логику, эпистемологию, лингвистику) и способствуют возникновению новых дисциплин – когнитивной психологии, компьютерной лингвистики. Поэтому анализ динамики области искусственного интеллекта имеет не просто историко-научный интерес, а является актуальным для изучения общих тенденций развития современной науки.

Актуальность настоящей работы связана также с «неклассическим» характером искусственного интеллекта как научно-технической дисциплины, что требует выработки нового методологического инструментария науковедческого анализа, отвечающего современному этапу взаимодействия науки и инженерии.

Цель и задачи работы. Целью диссертационного исследования является разработка методологической схемы науковедческого анализа современных неклассических научно-технических дисциплин на примере области искусственного интеллекта и применение данной схемы к исторической динамике этой области. Для достижения указанной цели решались следующие задачи:

- выявление основных методологических проблем науковедческого анализа области искусственного интеллекта;
- анализ применимости известных логико-методологических схем к области искусственного интеллекта;
- выработка адекватной науковедческой модели;
- историко-научное описание динамики области с использованием разработанной методологии;
- анализ особенностей рефлексии научного сообщества искусственного интеллекта и ее влияния на развитие исследований.

Теоретико-методологическая основа исследования. Диссертационная работа базируется на подходах современной логики, методологии и философии науки, общего науковедения. Используются, в частности, логико-методологические схемы развития науки, разработанные И. Лакатосом и Т. Куном [Лакатос, 1978; Кун, 1975], науковедческие модели и концепции, предложенные Б. Г. Гороховым, Н. И. Кузнецовой, А. А. Печенкиным, А. И. Ракитовым, М. А. Розовым [Горохов, 1984; Кузнецова, 1987; Печенкин, 1984; Ракитов и др., 1986; Ракитов, 1987; Розов, 1977, 1986, 1987].

В основу диссертационного исследования заложены представления о тесной взаимосвязи когнитивного и социального аспектов развития науки, о механизмах и факторах формирования научных областей и дисциплин (в том числе современных неклассических научно-технических дисциплин), о роли научного сообщества и его рефлексии в динамике развития исследований.

В качестве единицы описания развития области выбрано понятие

«исследовательской программы», являющееся модификацией «научной исследовательской программы» И. Лакатоса с учетом специфики искусственного интеллекта как неклассической научно-технической дисциплины. В соответствии с данной методологической схемой развитие области изучается в двух планах – как эволюция самих программ и как конкуренция между разными программами.

Соотношение научного и инженерного аспектов исследований по искусственному интеллекту анализируется на базе представлений о стирании граней между исследованием и проектированием в современных неклассических научно-технических дисциплинах и о рефлексивной симметрии предметного и рецептурного знания.

Научное сообщество рассматривается как рефлектирующая система, способная перестраивать свою деятельность в соответствии с выработанным самоописанием. Проводится четкое методологическое различие историко-научных событий, происходящих в деятельности ученых, и их отражения в рефлексии. Анализ рефлексии, ее особенностей и противоречий лежит в основе выработки адекватной логико-методологической схемы науковедческого анализа области.

Объект исследования, хронологические рамки и источники. В соответствии с данной методологией в качестве объекта анализа было выбрано единое, связанное тесными взаимными коммуникациями национальное научное сообщество. Поскольку сами исследования по искусственному интеллекту зародились в США, и это научное сообщество является и поныне общепризнанным лидером в данной области, то именно оно было выбрано как объект науковедческого анализа и историко-научного описания. В ряде случаев изложение дополнено материалами, касающимися исследований по искусственному интеллекту в СССР, Западной Европе и Японии.

Для исследования выбраны широкие хронологические рамки – с середины 1950-х годов по настоящее время. При этом история области искусственного интеллекта условно разделена на четыре этапа:

1950-е годы: зарождение исследовательских программ, первая конференция по искусственному интеллекту, организация лабораторий, первые результаты.

1960-е годы: начало чтения спецкурсов в университетах, бурный рост числа публикаций, расширение круга исследовательских программ.

1970-е годы: организация международных конференций, издание учебников, выпуск специализированных журналов: завершение становления искусственного интеллекта как научно-технической дисциплины.

1980-е годы: превращение систем искусственного интеллекта в коммерческий товар; новая фаза взаимодействия науки и инженерии.

На столь протяженном историческом отрезке были выбраны для рассмотрения лишь самые узловые моменты и события, важные с точки зрения развития и конкуренции исследовательских программ.

Круг первичных источников по истории искусственного интеллекта необычайно широк. Достаточно сказать, что библиография, составленная М. Минским в 1963 г. и дополненная в русском издании 1967 г., уже насчитывает 1185 названий. Более 1500 публикаций 1980-х гг. содержатся в библиографии трехтомного «Справочника по искусственному интеллекту» [Искусственный..., 1990]. О публикациях 1970-х гг. дают представление выпуски «Аннотированного указателя литературы по искусственному интеллекту» [Искусственный..., 1977-1987]. Подробная библиография содержится в англоязычных справочных изданиях [Encyclopedia..., 1987; Handbook..., 1981-1982], в

компьютерной базе данных [Scientific..., 1985].

В диссертационном исследовании использованы ключевые первичные публикации (статьи в журналах, труды конференций) и более широко – вторичные источники (обзоры, учебники, монографии, материалы философских дискуссий), в более явной форме выражающие рефлексию научного сообщества.

Новизна работы. Обзор историографии. Истории искусственного интеллекта посвящена не очень обширная литература, что выглядит вполне естественным, учитывая относительную молодость самой области. Непосредственно на эту тему написано несколько монографий [Бирюков, 1985; Поспелов Д. А., 1982; Boden, 1977; Haugeland, 1985; Johnson, 1986; McCorduck, 1979; Pratt, 1987] и ряд статей [Поспелов Г. С., 1980; Фейгенбаум, 1973; Шиклошши, 1973; Шрейдер, 1990; Cohen et al., 1988; Newell, 1983]. В то же время довольно велик круг работ, где встречаются отдельные важные исторические и методологические замечания.

Часть историографии посвящена в основном предыстории исследований по искусственному интеллекту, некоторые другие работы носят научно-популярный характер. Основной же массив исторических публикаций принадлежит либо самим специалистам по искусственному интеллекту, либо философам, интересующимся проблематикой мышления. Практически отсутствуют исследования философов науки, в которых анализировалась бы конкретная деятельность ученых, изучался искусственный интеллект как исследовательская область, а не сама проблема искусственного интеллекта. Наукометрические исследования, начатые Ж.-П. Куртьелем и Дж. Лоу с применением метода совместной встречаемости слов в литературе по искусственному интеллекту [Courtiel et al., 1989] находятся еще на ранней стадии развития.

Многие исторические и обзорные описания области искусственного интеллекта сфокусированы исключительно на успехах исследований (другие – лишь на неудачах); выбор материала явно субъективен и не дает полной картины развития области.

Важной особенностью большинства описаний области искусственного интеллекта является первоначальная разбивка исследований по подзадам (общение на естественном языке, игры, распознавание зрительных образов и т. д. ) и последующее изложение каждого раздела как независимого направления исследований [Минский, 1966; Поспелов Г. С. и др., 1974, 1975; Фейгенбаум, 1973]. При этом упускается из виду область искусственного интеллекта как целостный объект анализа. Дело в том, что один и тот же подход часто применяется при решении разных подзадач; действуют общие концептуальные установки, связывающие исследования по искусственному интеллекту в единый поток. Если описание «по подзадам» еще способно дать синхронический срез, «мгновенный портрет» дисциплины, то для отображения диахронического измерения, исторической динамики необходим иной подход.

Как заявил один из ведущих специалистов по искусственному интеллекту А. Ньюэлл, «в конце концов мы получим настоящую историю искусственного интеллекта, написанную с той объективностью, на которую способны историки науки. Но это время определенно еще не наступило» [Newell, 1983, p. 187]. Действительно, строго говоря, до настоящего времени область искусственного интеллекта не имеет ни подробного фактологического историко-научного описания, ни гносеологического описания, основанного на применении науковедческих аналитических средств. Более того, до сих пор не были сформулированы даже методологические основания для такого описания и анализа.

В данной диссертационной работе предлагается определенная методология и делается попытка ее применения к описанию ключевых моментов истории искусственного интеллекта и к построению гносеологического «портрета» этой области исследований.

На защиту выносятся следующие положения:

1. Область искусственного интеллекта как объект науковедческого исследования обладает следующими особенностями:

- результаты исследований по искусственному интеллекту «двойственны»: они могут рассматриваться и как научное знание, и как инженерный проект;
- предмет исследований имеет сложную двумерную структуру: это «интеллектуальное пространство» проблем – методов;
- граница области искусственного интеллекта с техникой программирования размыта, и элемент инженерии присутствует в любой теоретической разработке;
- быстрый темп исторической динамики области рождает определенное отставание рефлексии научного сообщества от деятельности отдельных ученых;
- мировоззренческий аспект проблемы искусственного интеллекта оказывает непосредственное воздействие на развитие конкретных научных исследований.

2. Логико-методологические схемы анализа развития науки, предложенные И. Лакатосом и Т. Куном, в их традиционной форме неприменимы к описанию области искусственного интеллекта.

3. Адекватная науковедческая модель для анализа области искусственного интеллекта включает модифицированную методологию «исследовательских программ» И. Лакатоса, где уточняются понятия «жесткого ядра», «защитного пояса» и «позитивной эвристики» и переопределяются критерии прогресса и регресса программ. Роль «жесткого ядра» играет исходная модель мышления, почерпнутая из смежной научной области, «защитного пояса» – ее конкретная реализация в виде компьютерной программы, а «позитивная эвристика» включает круг интеллектуальных проблем из той же смежной области и набор релевантных программных средств.

4. Критерий прогресса исследовательских программ связан с объемом «интеллектуального пространства», в котором «движутся» программы. При этом вырождается та исследовательская программа, что охватывает уже известные проблемы и методы, а прогрессирует программа, расширяющая само «интеллектуальное пространство» за счет новых проблем и методов.

5. В области искусственного интеллекта функционируют, конкурируя, пять исследовательских программ:

- *логистическая* (использующая исходную модель «мышление – это логический вывод»);
- *эвристическая* (основанная на психологической модели мышления как поиска пути в лабиринте);
- *коннекционистская* (использующая нейрофизиологическую гетерархическую модель нейронной сети);
- *эволюционная* (базирующаяся на биологическом механизме случайных мутаций и естественного отбора);
- *когнитивная* (использующая эпистемологическую модель «мышление – это

познание»).

6. В исторической динамике взаимодействие программ развивалось по следующей схеме:

В 1950-е годы, когда возникли первые четыре исследовательские программы, разнообразие поставленных ими интеллектуальных задач было столь велико, что все программы прогрессировали, не испытывая конкуренции.

В 1960-е годы «традиционные» исследовательские программы испытали кризис при переходе к широкому классу реальных, а не лабораторных задач. Явно прогрессировала лишь новая, возникшая в эти годы когнитивная программа.

В 1970-е годы резко проявилось расхождение научных и инженерных критериев оценки. Если в инженерной области большего успеха добилась эвристическая программа, то в научной – когнитивная.

В 1980-е годы логистическая, эвристическая и когнитивная программы стали в основном поставлять инженерные изделия на коммерческий рынок, а научный прогресс ждал коннекционистскую программу.

7. Рефлексия специалиста по искусственному интеллекту обладает следующими особенностями:

- ученый, принадлежащий к неклассической научно-технической дисциплине, по специфике своей работы обязан занимать методологическую, или «мета-рефлексивную» позицию;
- рефлексия специалиста входит в содержание его собственной деятельности: он анализирует свою работу, строя модели научного познания и рефлексии для систем искусственного интеллекта.

8. Сложная, противоречивая картина рефлексии научного сообщества искусственного интеллекта определяется следующими факторами:

- комплексной структурой предмета исследования («интеллектуального пространства»);
- полифункциональностью самих исследований;
- разнообразием процедур рефлексии.

9. Дискуссия по проблеме «Может ли машина мыслить?» оказывает влияние на деятельность ученых в нескольких аспектах:

- мировоззренческом: ставит моральные вопросы об этичности передачи машинам некоторых интеллектуальных функций и трактовки самого человека как машины, перерабатывающей информацию;
- практическом: во многом определяет выбор терминологии, решаемых задач, используемых методов и критериев оценки систем;
- гносеологическом: стимулирует поиски неантропоморфного определения мышления.

Научно-практическая значимость работы определяется тем, что в ней впервые разработана специализированная методологическая схема науковедческого анализа области искусственного интеллекта и дана целостная картина исторической динамики данной области.

Результаты диссертационного исследования могут быть использованы для научно обоснованной оценки современного состояния и перспектив развития области искусственного интеллекта, в последующих исследованиях ее истории. В обобщенном виде предложенная методологическая схема может применяться при анализе других

неклассических научно-технических дисциплин, использоваться в работах по философии и методологии науки. Материалы диссертации могут служить источником при разработке и чтении спецкурсов по философским проблемам современного естествознания, науковедению, истории науки.

Структура диссертации обусловлена целями и методологией исследования. Работа состоит из введения, трех глав, заключения и списка литературы.

В первой главе – «Методология науковедческого анализа области искусственного интеллекта» – дается характеристика особенностей области искусственного интеллекта и методологических проблем ее анализа. Известная методологическая схема описания развития науки, данная И. Лакатосом, корректируется с учетом специфики исследуемой области.

Во второй главе – «Историческая динамика области искусственного интеллекта» – предложенная схема развития и конкуренции исследовательских программ накладывается на исторический материал. Узловые моменты в развитии исследований по искусственному интеллекту получают свое объяснение в рамках данной методологической схемы.

В третьей главе – «Роль рефлексии ученых в развитии исследований по искусственному интеллекту» – рассматриваются выработанные в области искусственного интеллекта модели научного познания и рефлексии и исследуется их взаимосвязь с рефлексией ученых над собственной деятельностью. Анализируется влияние на деятельность ученых дискуссии по проблеме «Может ли машина мыслить?» в ее мировоззренческом, практическом и гносеологическом аспектах.

Список литературы включает 220 названий.

Апробация работы. Основные положения и результаты диссертации были апробированы на конференциях «Мировоззренческие и методологические проблемы компьютеризации современной науки» (Обнинск, 1986), «Методологические проблемы взаимосвязи фундаментальных исследований и разработки интенсивных технологий» (Обнинск, 1987), Всесоюзном семинаре «Формы представления знаний и творческое мышление» (Новосибирск, 1989), IV чтениях памяти акад. В. М. Кедрова (Одесса, 1990), XXXII и XXXIII научных конференциях аспирантов и молодых специалистов по истории естествознания и техники (Москва, 1990, 1991). Опубликованы тезисы выступлений [Герович, 1989, 1990а, 1990в] и препринт [Герович, 1990б].



## **Глава 1. МЕТОДОЛОГИЯ НАУКОВЕДЧЕСКОГО АНАЛИЗА ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

### **1. 1. Область искусственного интеллекта как объект науковедческого анализа**

#### **1. 1. 1. Постановка проблемы**

Исследования по искусственному интеллекту представляют собой в высшей степени необычный и интересный для науковедческого анализа объект. Характеризуя становление информационной науки, сердцевиной которой является искусственный интеллект, А. И. Ракитов писал: "В наше время можно наблюдать, как различные разделы информатики, машинной лингвистики, когнитивной психологии, эпистемологии, математической логики, теории автоматов, различных разделов физики, математики и т. д., развивавшиеся ранее в рамках изолированных научных дисциплин, выделяются из их состава и, постепенно трансформируясь, вступая в тесные взаимодействия, превращаются в совершенно новую систему знаний, беспрецедентную с точки зрения всей предшествующей истории науки. Я даже склонен думать, что благодаря этому влиянию мы в недалеком будущем сможем стать современниками и очевидцами фундаментальной реорганизации всего научного знания. Именно это обстоятельство должно сейчас находиться в центре внимания историков и методологов науки, именно оно может и должно дать материал для построения новых широких науковедческих и историко-научных концепций функционирования и развития науки в условиях компьютерной революции" [Ракитов, 1987, с. 38-39].

В данной главе мы остановимся на тех проблемах, с которыми сталкивается науковед, пытаясь применить к анализу области искусственного интеллекта свой обычный инструментарий. На наш взгляд, корень этих проблем – в специфике изучаемого объекта, и для успешного исследования необходимо развитие самих аналитических средств, введение новых моделей.

Внешне история искусственного интеллекта выглядит вполне благополучно. За точку отсчета единодушно принимается 1956 год, когда в рамках Летнего исследовательского проекта в Дартмутском колледже (США) собралось несколько ученых, составивших костяк нового направления. Большой научный и общественный резонанс, который получили первые работы по искусственному интеллекту, привлек к ним повышенное внимание и способствовал притоку ученых из смежных областей.

В 1957 г. открылась первая лаборатория искусственного интеллекта в Массачусетском технологическом институте, в 1963 – в Стенфордском университете. В конце 1960-х – начале 1970-х гг. появились первые учебники по искусственному интеллекту, началась целенаправленная подготовка кадров в университетах.

В те же годы складываются устойчивые формы коммуникации, связывающие ученых в научное сообщество: с 1969 г. каждые два года созываются Международные объединенные конференции по искусственному интеллекту, с 1970 г. издается международный журнал "Artificial Intelligence". Впоследствии число конференций и журналов многократно умножилось. В настоящее время за искусственным интеллектом закрепились репутация переднего края компьютерной науки.

В то же время для многих ведущих ученых вопрос о предмете и целях их исследований был и остается проблематичным. Видный специалист по искусственному

интеллекту, автор одного из первых учебников, профессор факультета психологии университета в Сиэтле Эрл Хант в 1975 г. так описывал состояние дел: "Если бы физики или химики взялись дать абстрактные определения своих областей знания, то вы скорее всего не нашли бы разногласий ни среди тех, ни среди других. Вряд ли бы обнаружилось такое единодушие, если бы пришлось собрать вместе разных ученых, занимающихся искусственным интеллектом... Оказалось, что существует содержательная область знаний, общие принципы которой трудно выделить. Проблема, по-видимому, заключается в определении понятия интеллекта" [Хант, 1978, с. 11-12]. Даже сам "отец искусственного интеллекта", считающийся автором этого термина, организатор Дартмутского проекта, глава лаборатории искусственного интеллекта в Стенфордском университете Джон Маккарти затруднялся в ответе на вопрос, что же такое "искусственный интеллект": "Я думаю, что определение понятия интеллекта – это составная часть теории интеллекта, а я совсем не готов предложить такую теорию" [Дискуссии..., 1978, с. 38]. В том же 1975 году ту же проблему, что и Э. Хант, рассматривали ведущие отечественные специалисты Г. С. Поспелов и Д. А. Поспелов. Среди ученых, писали они, "нет единого мнения о предмете их занятий. Что понимать под искусственным интеллектом, каково его отношение к интеллекту человека, где проходят границы, отделяющие интеллектуальную деятельность от неинтеллектуальной? На эти вопросы пока нельзя ответить со сколько-нибудь приемлемой полнотой" [Поспелов Г. С. и др., 1975, с. 26]. В последнее время число таких высказываний несколько убавилось, но это произошло не из-за прояснения ситуации, просто научное сообщество, по-видимому, смирилось с такой имманентной неопределенностью.

Конечно, область искусственного интеллекта не осталась вовсе без характеристики; ученые вынуждены были принять те или иные рабочие формулировки. В определениях не было недостатка, можно даже сказать, что их накопилось слишком много, ибо выявилось больше противоречий между ними, чем сходств. Сам по себе этот материал – вербализованная рефлексия ученых – является ценнейшим свидетельством об основном объекте нашего анализа – исследовательской области искусственного интеллекта. Анализируя корни противоречий в рефлексии, выявляя специфику исследований по искусственному интеллекту, мы поймем, что эта область и не могла иметь четкого и стройного самоописания, что эта неопределенность, как ни странно, действительно имманентна.

### **1. 1. 2. Искусственный интеллект: инженерия или наука?**

Традиционно инженерия и наука различаются по целям, которые при этом преследуются. Цель инженера – разработка проекта, критерии оценки которого – реализуемость и эффективность. Ученый же добывает знание, оцениваемое на истинность или ложь, строит модели, заботясь об их адекватности исследуемым объектам. К знанию бесполезно подходить с меркой эффективности, так же как к проекту – с попыткой доказать его "истинность". Если данная ситуация представляется совершенно ясной для "образцовых" научных дисциплин типа физики, то в случае компьютерной науки возникают неожиданные трудности.

Модели мышления, которые строятся в искусственном интеллекте, реализуются в виде компьютерных систем, выполняющих соответствующие интеллектуальные функции. Чем является такая система – знанием о том как мыслит человек (или некоторое

абстрактное разумное существо) или же проектом "интеллектуальной машины"? Оценивать этот продукт по критерию "истинно–ложно" или по критерию "эффективно–неэффективно"? От ответа на этот вопрос зависит квалификация искусственного интеллекта как научной или инженерной дисциплины и, соответственно, выбор подходящего инструментария науковедческого анализа.

Фундаментальная проблема заключается в том, что в данном случае ответ на этот вопрос субъективен, он зависит от "внутренней" цели, которую ставит перед собой каждый конкретный исследователь. Ибо объективно система искусственного интеллекта может выступать и в качестве научного продукта, и как инженерная разработка.

Коль скоро мнение о характере исследований по искусственному интеллекту субъективно, то сколько ученых – столько и мнений. Мы соберем это разнообразие в несколько групп, объединяющих близкие точки зрения, и попытаемся найти в природе исследований конкретные причины, обуславливающие именно такой набор мнений. Явно выделяется пять основных точек зрения на данную проблему:

- 1) Искусственный интеллект – набор различных научных и инженерных дисциплин;
- 2) Искусственный интеллект – инженерия;
- 3) Искусственный интеллект – наука;
- 4) Искусственный интеллект – и наука, и инженерия одновременно;
- 5) Искусственный интеллект – не наука и не инженерия.

Оказывается, что такой немислимый по противоречивости спектр мнений о характере исследований целиком заполняется высказываниями ученых. Проиллюстрируем это на конкретных примерах.

Первая точка зрения наиболее проста и состоит в попытке обойти вышеозначенные трудности. Научная и инженерная компоненты одних и тех же исследований при этом разделяются и как бы обособляются в самостоятельные дисциплины. Э. Рич, например, считает, что в искусственный интеллект входят "как научные, так и инженерные дисциплины" [Rich, 1987, p. 9]. Более подробно ту же мысль излагает Дж. Слэйгл: "Исследования в области 'искусственного интеллекта'... имеют двоякую цель: выяснение сущности естественного (человеческого) интеллекта и использование машинного интеллекта для приобретения новых знаний и решения интеллектуально-трудных задач... Первую цель преследует психолог. Он конструирует в виде программы для вычислительной машины модель определенного поведения, рассматриваемого как человеческий интеллект... Исследователь, ставящий перед собой вторую цель, заинтересован лишь в синтезе интеллектуального поведения и не заботится о том, использует ли вычислительная машина методы, которыми пользуются люди" [Слэйгл, 1973, с. 13-14]. Интересно, что такой же точки зрения придерживается и известный английский критик искусственного интеллекта сэр Джеймс Лайтхилл. Правда, он делает отсюда вывод о том, что все, происходящее в области искусственного интеллекта, с таким же успехом могло бы протекать отдельно в исследованиях по вычислительной технике, теории управления и психологии, и нет никакой необходимости в "перемычке", связывающей ряд уже сложившихся дисциплин.

Вторая группа высказываний принадлежит специалистам по искусственному интеллекту, считающим себя чистыми инженерами. Один из ведущих сотрудников группы искусственного интеллекта Стенфордского исследовательского института Нильс Нильсон высказался на этот счет совершенно определенно: "Моя точка зрения состоит в том, что искусственный интеллект представляет собой ... инженерную дисциплину,

поскольку его первоначальной целью является создание конструкций" [Нильсон, 1973, с. 7]. Аналогично мнение видного отечественного специалиста Э. В. Попова: "Искусственный интеллект – инженерная дисциплина, предназначенная для разработки машин и программ, способных к деятельности, называемой интеллектуальной, если она выполняется людьми" [Искусственный..., 1979, № 4, с. 68]. Следует отметить, что эта точка зрения, особенно в последнее время, заметно количественно преобладает. Достаточно привести сделанное в 1984 году горькое признание представителя оппозирующего лагеря, главы школы искусственного интеллекта в Йельском университете Роджера Шенка: "Я всегда больше ориентировался на моделирование человеческого разума, чем другие; меня занимает воспроизведение процессов человеческого мышления. Я больше интересуюсь людьми, чем машинами, и это ставит меня вне большинства сообщества искусственного интеллекта [Цит. по: Johnson, 1986, p. 179].

В третьей группе мнений наблюдается внутренний раскол по вопросу о том, исследуется ли в искусственном интеллекте сугубо человеческий интеллект или же выявляются некие общие принципы мышления, не обязательно связанные с человеческим субъектом. Мнение первой "подгруппы" выразил, в частности, Р. Шенк: "Область науки, называемая "Искусственный интеллект", занимается изучением самых таинственных вопросов человеческого существования. Какова природа мышления? Какие процессы происходят в нашем организме, когда мы думаем, чувствуем, видим, понимаем? Возможно ли в принципе понять, как работает наш мозг?" [Шенк и др., 1987, с. 15]. Эту же точку зрения разделяют основатели лаборатории искусственного интеллекта в Университете Карнеги-Меллона, одни из "пионеров" области, участники Дартмутского проекта Алан Ньюэлл и Герберт Саймон. По мнению Ньюэлла, искусственный интеллект является, по сути дела, теоретической психологией, подобно тому, как математическая теория дифференциальных уравнений играет роль теоретической физики. "Теории когнитивной психологии должны быть выражены в виде систем искусственного интеллекта" [Newel 1, 1970, p. 368]. Более развернуто формулирует свое кредо Саймон: "Можно написать программы, которые, используя процессы манипуляции нечисловыми символами, решают задачи, требующие от человека размышлений и обучения. Подобные программы могут быть названы теориями, в буквальном смысле этого слова, соответствующих мыслительных процессов. Эти теории допускают проверку многими различными способами. Среди этих способов можно отметить сравнение символического поведения вычислительной машины, выполняющей данную программу, с символическим поведением человека, когда и машина, и человек решают одну и ту же задачу" [Дискуссии..., 1970, с. 22].

Для второй "подгруппы" характерно высказывание философа М. Рингля, считающего целью искусственного интеллекта как науки создание общей теории интеллекта, включающей в себя теорию человеческого интеллекта [Ringle, 1979, p. 13]. Нельзя не отметить, что эта точка зрения приобретает все большую привлекательность для специалистов, осознающих ограниченность чисто инженерного подхода, но избегающих "психологизма", связанного с моделированием сугубо человеческого процесса мышления. Еще в 1971 г. Н. Нильсон уверенно заявлял, что "в поисках теории искусственного интеллекта смысла не больше, чем в поисках, скажем, теории гражданского строительства. Вместо единой общей теории имеется ряд теоретических дисциплин, которые сюда относятся и которые должны изучаться теми, кто выбирает

искусственный интеллект своей специальностью. К таким дисциплинам относятся математическая логика, структурная лингвистика, теория вычислений, теория информационных структур, теория управления, статистическая теория классификации, теория графов и теория эвристического поиска" [Нильсон, 1973, с. 7]. Однако уже в 1974 г., на IV конгрессе Международной федерации информационной обработки (IFIP) в Стокгольме, он признал существование собственной "территории" искусственного интеллекта: "В фундаментальном смысле искусственный интеллект есть наука о знании, т.е. наука о том, как представлять, получать и использовать знание" [Nilsson, 1974, p. 7]. Более того, он объявил искусственный интеллект наукой, а не инженерией, как раньше! Где произошел этот сдвиг: в реальной истории дисциплины или только в рефлексии отдельного исследователя? Мы еще вернемся к этому вопросу, ибо решить его можно, лишь выработав адекватную модель искусственного интеллекта как объекта науковедческого анализа. Тем временем продолжим рассмотрение точек зрения на характер исследований в данной области.

Согласно четвертой группе мнений, научная и инженерная компоненты исследований присутствуют одновременно и неотделимы друг от друга. Американский исследователь науки Джордж Джонсон так характеризует особенности области искусственного интеллекта: "Разделение между чистым и не прикладным – не столько раскол, сколько полярность – напряжение, которое движет поле. Наука и технология всегда играли друг против друга. Теории ведут к приложениям, чьи недостатки предлагают направления дальнейших исследований. Старые науки типа физики имели время разделиться на теоретическую и прикладную части. Искусственный интеллект же возник только сейчас, и он так быстро развивается, что граница между наукой и инженерией часто расплывается" [Johnson, 1986, p. 28]. Такая ситуация резко осложняет проблему оценки систем искусственного интеллекта. Директор лаборатории интеллектуальных систем компании "Ксерокс" Дж. С. Браун также сетовал на то, что из-за "молодости" исследовательской области нельзя оценить, является ли данная программа теорией разума или только удачным инженерным творением. Если система демонстрирует интеллектуальное поведение, то это происходит из-за адекватности научной теории этого поведения или благодаря искусству программирования? Программистская техника может быть достаточной для генерирования определенного поведения, но это не означает, что интеллектуальная система должна с необходимостью работать именно таким образом [Ibid., p. 241-242]. Данное понимание взаимодействия научной и инженерной компонент исходит из представления об иерархической структуре исследования: компоненты работают вместе, и каждая выполняет свою часть общего дела. Та же идея связи науки и инженерии у Р. Бенерджи, автора первого учебника по искусственному интеллекту, выглядит совсем иначе. Анализируя отдельные работы, пишет он, "в каждом конкретном случае трудно сказать, является ли это попыткой моделирования мыслительного процесса или попыткой решения некоторой задачи" [Бенерджи, 1972, с. 17]. При этом одну и ту же систему искусственного интеллекта можно рассматривать, то как плод инженерии, то как научный продукт. Наука и инженерия выглядят как две стороны одной медали, одного процесса, а не его составные части.

Последняя группа высказываний о природе искусственного интеллекта обязана своим рождением в основном профессиональным критикам этой области, среди которых особой непримиримостью выделяется американский философ Х. Дрейфус. Еще в 1965 г. он написал работу "Алхимия и искусственный интеллект", позднее переработанную в

книгу "Чего не могут вычислительные машины?" [Дрейфус, 1978]. Дрейфус проводит прямую аналогию между областью искусственного интеллекта и алхимией, между поисками моделей мышления и философского камня, упрекая исследователей искусственного интеллекта в игнорировании научных критериев проверки результатов. С ним солидаризуется профессор факультета психологии Массачусетского технологического института Дж. Вейценбаум, автор нашумевшей диалоговой системы "Элиза", ловко создававшей иллюзию "понимания" естественного языка. Подозревая, по-видимому, наличие подобных трюков и в других системах искусственного интеллекта, он пишет: "Ньюэлл, Саймон, Шенк и Виноград просто заблуждаются относительно природы тех задач, "решением" которых они, по их мнению, занимаются. Они... выдвигают "общие теории", являющиеся, в конечном счете, пустыми эвристическими лозунгами, а затем провозглашают эти "теории" подтвержденными после создания моделей, которые действительно решают некоторые задачи, но таким образом, что исключается возможность выяснения чего-нибудь относительно общих принципов" [Вейценбаум, 1982, с. 255]. Надо отметить, что хотя Дрейфус и называет искусственный интеллект "наименее самокритичной областью во всем конгломерате естественных наук", здесь все же есть ученые, способные отнестись к своей работе не менее взыскательно, чем критики "извне". Директор лаборатории искусственного интеллекта Массачусетского технологического института Патрик Уинстон тоже считает, что уход от состояния алхимии еще только предстоит: "Искусственный интеллект – это искусство, собирающееся стать наукой, с очагами инженерии то тут, то там" [Цит. по: Johnson, 1986, p. 241]. Итак, коротко говоря, пятая точка зрения состоит в том, что искусственный интеллект – это интеллектуальное искусство.

Наконец, приведем результаты опроса специалистов по искусственному интеллекту, проведенного в 1978 г. Р. Бречменом и Е. Смитом. На вопрос о цели их собственных исследований 25% опрошенных назвали проектирование интеллектуальных систем, 45% – исследование теоретических проблем, а 25% – и то, и другое [Special..., 1980, p. 51-52].

### **1. 1. 3. Гносеологическая характеристика области искусственного интеллекта**

Самоописание области искусственного интеллекта, продемонстрированное выше, пестро и разноречиво. Научное сообщество расколото на множество несогласных друг с другом групп. Чем объяснить такое разнообразие?

Может быть, гетерогенностью самой области, когда одни исследователи действительно занимаются только наукой, а другие – только инженерией, и речь просто идет о разных вещах? Но все дело в том, что мнения расходятся как раз относительно одного и того же предмета, одних и тех же исследований. Кроме того, нет жесткости в позициях сторон – нередко один специалист мог высказываться то в научном, то в инженерном духе.

Тогда, вероятно, играет роль смена характера исследований во времени: научные знания претворяются в инженерные проекты или, наоборот, работа технических устройств дает материал для последующих теоретических обобщений? Безусловно, эти процессы имеют место, но они не объясняют синхронического разнообразия мнений – ведь инженерия и наука не сменяют друг друга в истории искусственного интеллекта, а присутствуют с самого начала на всех этапах развития области.

Кто же из высказавшихся выше ученых прав? Если предположить, что какая-либо

одна из точек зрения верна, а остальные нет, то большинство ученых автоматически окажутся пребывающими в заблуждении относительно того, чем они занимаются на работе. Не будем торопиться объяснять разнообразие мнений ошибками и неправильными представлениями. Попробуем понять, каков сам объект, вызвавший столь разноречивые о себе представления у людей, работающих в соседних лабораториях.

Гносеологический "портрет" области искусственного интеллекта, на наш взгляд, обладает следующими чертами;

1) Двойственность исследований. Несомненным продуктом искусственного интеллекта являются компьютерные системы, выполняющие определенные интеллектуальные функции. Этот объект оказывается "двуликим": его можно рассматривать и как научное знание о том, как человеческое мышление осуществляет данную функцию, и как инженерный продукт, проект устройства, решающего определенную задачу. Субъективная цель, которую ставит перед собой создатель конкретной системы, не играет при этом роли: потенциальная включенность в деятельность двух видов является свойством самого объекта. Расхождения в рефлексии ученых объясняются просто: суждения о научном или инженерном характере работ по искусственному интеллекту не противоречат друг другу, но лишь сделаны с разных позиций.

Как точно отмечает В. Г. Горохов, представитель современной научно-технической дисциплины выступает сразу в нескольких ролях: исследователя (следующего нормам научной теоретической деятельности), инженера-проектировщика (создающего проект), методиста и других [Горохов, 1984, с. 67]. При этом сам ученый часто осознает себя лишь в одной роли и, не учитывая специфики своей области, стремится реализовать классический идеал построения научного знания, что оказывается невозможным.

Это приводит к неопределенности самооценки и порождает как методологические, так и чисто психологические проблемы. "Представители этих дисциплин не могут себя отнести ни к теоретикам, ни к практикам, ни к исследователям, ни к проектировщикам и т.д. и тем самым ощущают себя за пределами традиционной организации науки и инженерии" [там же, с. 80].

2) Сложность предмета исследования. К проблеме искусственного интеллекта можно подойти двояко: создать компьютерную систему, реализующую определенный метод моделирования (например, логический), и проверять ее на последовательности интеллектуальных проблем, или же начать создавать систему "под проблему" (скажем, игру в шахматы), пробуя различные методы для ее решения. В "интеллектуальном пространстве" можно двигаться как "горизонтально" – по проблемам (при заданном методе), так и "вертикально" – по методам (при фиксированной проблеме). Образующаяся при этом "решетка" и составляет предмет искусственного интеллекта – "пространство" мышления.

Восстановить это пространство в полном объеме – самостоятельная задача. Одну из первых попыток такого рода предпринял Л. Т. Кузин, составивший таблицу, по строкам которой располагаются методы моделирования, а по столбцам – интеллектуальные проблемы [Кузин, 1974]. Среди методов присутствуют: человеко-машинный, лингвистический, эвристический, логико-алгебраический, статистический, оптимизационный и бионический. К интеллектуальным проблемам относятся: распознавание образов, индуктивный вывод, формирование понятий, доказательство теорем, ответы на вопросы, игры, решение задач. Эти списки могут быть уточнены и

расширены, но для нас важно отметить саму структуру пространства, в котором разворачивается поиск искусственного интеллекта.

3) Размытость границы. Создание системы искусственного интеллекта – процесс, требующий не только построения теоретической модели мышления, но и ее программной реализации. Сдвигаясь "вниз", к непосредственным исполнителям – программистам, работа на каком-то этапе теряет черты теории и становится чисто техническим делом. Но строго определить эту границу невозможно. Ограничения или, наоборот, дополнительные возможности, которые дает современное программирование, постоянно видоизменяют теоретическую модель. Р. Шенк верно отмечает, что "после того, как идеи искусственного интеллекта превращаются в прикладную систему, они в каком-то смысле переходят рамки искусственного интеллекта" [Шенк и др., 1987, с. 24]. Сложность, однако, заключается в том, что пока эти идеи не воплощены в системе, пока не поработала инженерия, нельзя оценить и адекватность данной модели мышления, нельзя рассматривать ее как научное знание.

4) Трансформация области в исторической динамике. Бурное развитие области искусственного интеллекта, быстрая проверка гипотез, постоянный приток новых людей и свежих идей сделали гносеологическую картину происходящего чрезвычайно изменчивой.

В одни периоды несколько преобладали научные постановки вопроса, в другие по разным причинам усиливался инженерный подход. Это не могло не отразиться на рефлексии научного сообщества, которая дифференцировалась в соответствии с тем, кто успевал (или нет) следить за всеми новациями. Одни исследователи меняли взгляды, как Н. Нильсон (см. выше его мнения по проблеме "инженерия–наука"), другие сохраняли свои представления прежними, отмечая новое как ересь. Конкретный исторический анализ, который будет дан во второй главе, позволит детальнее рассмотреть и изменения самой области искусственного интеллекта, и перемены в представлении научного сообщества о ней.

5) Мировоззренческий смысл. Э. Хант недаром предостерегал читателей своей книги: "Любые обсуждения искусственного интеллекта должны проходить на очень техническом уровне, иначе они окажутся 'в тисках' вопроса о том, как изучение думающих машин должно или может влиять на наши философские и психологические представления" [Хант, 1978, с. 524]. Сама идея создания искусственного интеллекта, последующие попытки, успехи и неудачи вызвали невероятный всплеск эмоций журналистов, скептицизма психологов, инженерного оптимизма и философской критики. Если в предложенной В. Г. Гороховым схеме "нормального" развития научно-технической дисциплины научно-популярные издания должны появиться лишь на третьем этапе (через 10-20 лет после зарождения нового направления), то идеи искусственного интеллекта общественность начала обсуждать едва ли не раньше, чем ученые-профессионалы. Неиссякаемая дискуссия на тему "Может ли машина мыслить?" окружила шлейфом домыслов, неоправданных восторгов и необоснованной критики всякую систему искусственного интеллекта, получившую хоть какую-нибудь известность. Реальная научная и инженерная ценность работы при этом совершенно терялась. Необходимость постоянно отвечать на мировоззренческие вопросы, возникающие в связи с их работами, заставляла ученых по-иному смотреть на свою деятельность, и, как будет показано в третьей главе данного исследования, отражалась на когнитивной сфере. Без постоянного мировоззренческого "пресса" гносеологическая картина области была бы неполна.



Перечисленные черты характеризуют область искусственного интеллекта как весьма нестандартный объект науковедения. Эта характеристика решает одну проблему – объясняет разнообразие в рефлексии научного сообщества, но ставит другую – проблему выбора средств дальнейшего науковедческого анализа.

## **1. 2. Анализ применимости различных схем развития науки к истории искусственного интеллекта**

### **1. 2. 1. Обзор существующих подходов**

Современная философия науки, со всем богатством выработанных в ее рамках моделей развития науки, выросла из остро осознанного в 1930-е годы разрыва между логикой и историей науки. Логика науки анализировала научную "норму", процесс познания, "очищенный" от исторических условностей, а история описывала деятельность ученых во всей конкретности, но без логико-методологического стержня, скрепляющего последовательность разнообразных событий в причинно-следственную цепь. Задача философии науки – совмещение логической и исторической картин науки, создание моделей, объясняющих динамику когнитивной и социальной сфер, механизм появления нового в науке. Неопозитивистская логика науки была способна лишь анализировать структуру сложившихся научных теорий, но отступала перед любым серьезным научным движением, преобразованием наличной системы знаний. История же науки могла только зафиксировать это движение, но не объяснить его.

Проблема была сфокусирована в самом названии книги К. Поппера "Логика научного открытия" [Поппер, 1983], впервые вышедшей на немецком языке в 1934 г. Отыскать логику в развитии науки, объяснить не только статику, но и динамику – такую цель поставила перед собой постпозитивистская философия науки. Предложенный Поппером механизм "предположений и опровержений" опирался на ряд принципов, управляющих научным исследованием – выдвижение наиболее смелых, маловероятных гипотез, критицизм, фальсифицируемость (взамен позитивистской верификации), фаллибилизм (вместо кумулятивности). Не останавливаясь подробно на этих хорошо известных принципах, мы лишь укажем на концепцию Поппера как отправную точку для формирования двух других чаще используемых и активно обсуждаемых в контексте истории искусственного интеллекта логико-методологических схем, принадлежащих Томасу Куну и Имре Лакатосу.

Т. Кун обратил внимание на то, что модель Поппера описывает лишь кратковременные революционные периоды в развитии науки, в остальное же время действует скорее не принцип критицизма, а подчинение догме, господствующей в научном сообществе. В этот период, названный Куном "нормальной наукой", ученые, как правило, заняты не выдвижением смелых гипотез, а решением традиционных "головоломок" по имеющимся образцам. Важнейшая мысль Куна – о научном сообществе как подлинном субъекте науки. Взамен одинокого исследователя, руководствующегося лишь собственным разумом, на авансцену выступает группа ученых, для которой то или иное методологическое решение (например, принятие некоторой теории) определяется уже не только рациональными, но и социальными и психологическими факторами.

Концепция Куна оформилась в виде логико-методологической схемы, в которой последовательно, в результате научных революций, сменяются "парадигмы" – устойчивые

системы научных ценностей, образцов деятельности, метафизических убеждений и формализмов. Принятые научным сообществом в качестве незыблемой основы, они определяют весь спектр допустимой деятельности на период "нормальной науки". Парадигма задает и типы выдвигаемых задач, и образцы их решения, и критерии оценки научных результатов.

С момента первого издания в 1962 г. книги Куна "Структура научных революций" [Кун, 1975] эта концепция быстро стала популярной не только у коллег-философов науки, но и в среде ученых из частных дисциплин. Слова "парадигма", "парадигмальный" вошли обиход профессиональных научных сообществ, потеряв при этом, конечно, многие тонкости, заложенные Куном в его концепцию. Тем не менее, ученые получили язык, на котором могли выразить рефлексивное осознание сдвигов, происходящих в областях их исследований.

Кун считал, что решения научного сообщества выглядят рациональными лишь внешне, на деле же определяющую роль играет социально-психологический пресс господствующей догмы. В его концепции, однако, не был прояснен механизм возникновения новой парадигмы и вытеснения ею старой, а также источник разногласий среди ученых, принадлежащих к единой парадигме, по многим частным вопросам. Ученик Поппера И. Лакатос попытался преодолеть эти трудности, "реабилитировав" рационализм и попытавшись под слоем социально-психологических факторов отыскать еще более глубокие рациональные основания.

Лакатос согласился с Куном в критике попперовского императива "опровергнув – отвергай", но считал при этом, что дело не в настроении научного сообщества, а в наличии или отсутствии лучшей теории, которая могла бы заменить опровергнутую. Отсутствие такой теории заставляет ученых не придавать решающего значения аномалиям и лишь подправлять старую теорию, "ремонтировать" сложившуюся систему знаний. Если же подходящая новая теория имеется в наличии, то она вытесняет старую на вполне рациональных, логических основаниях.

Лакатос уделил большее внимание научной мотивации – тому, что побуждает ученых искать лучшие теории. "Парадигмального" ученого в схеме Куна практически ничто не способно вывести из равновесия. Ученый у Лакатоса испытывает постоянную конкуренцию со стороны коллег, придерживающихся иной "исследовательской программы".

Исследовательская программа – центральное понятие в схеме Лакатоса – представляет собой последовательность теорий, имеющих общее "жесткое ядро" и различающихся "защитными поясами". "Жесткое ядро" содержит наиболее фундаментальные положения, не изменяющиеся в процессе развития научной исследовательской программы. "Защитный пояс" состоит из дополнительных гипотез, частных уточнений и модификаций теории, позволяющих учесть аномалии и предохранить от критики "жесткое ядро". Важно отметить, что каждая исследовательская программа имеет собственный механизм выдвижения новых проблем и "защитных поясов" – "позитивную эвристику".

Конкуренция программ заставляет ученых соревноваться в широте круга решаемых задач, быстроте учета аномалий и стройности выдвигаемых теорий. Для оценки результатов этой борьбы Лакатос предложил критерий "прогрессивности сдвига проблем": "Исследовательская программа считается прогрессирующей тогда, когда ее теоретический рост предвосхищает ее эмпирический рост, т. е. когда она с некоторым

успехом может предсказывать новые факты ('прогрессивный сдвиг проблем'); программа регрессирует, если ее теоретический рост отстает от ее эмпирического роста, т. е. когда она дает только запоздалые объяснения либо случайных открытий, либо фактов, превосходяемых и открываемых конкурирующей программой ('регрессивный сдвиг проблем'). Если исследовательская программа прогрессивно объясняет больше, нежели конкурирующая, то она 'вытесняет' ее, и эта конкурирующая программа может быть устранена (или, если угодно, 'отложена')" [Лакатос, 1978, с. 219-220].

Логико-методологическая схема Лакатоса сочетает элементы относительной устойчивости ("жесткое ядро") с мощным внутренним генератором изменений (конкуренция программ). Некоторая парадоксальность (и в то же время, на наш взгляд, большое достоинство) этой модели заключается в отсутствии в ней центральных событий всей прежней историографии науки – научных революций. Одна исследовательская программа редко одерживает окончательную победу над другой. А. А. Печенкин справедливо обращает внимание на роль социальных факторов в конкуренции программ: "В принципе критерием научности исследовательской программы является лишь осведомленность научной общественности обо всех успехах и неудачах этой программы. Научная исследовательская программа может неопределенно долго находиться в состоянии регрессивного сдвига, поддерживаемая лишь упорством сторонников ее жесткого ядра" [Печенкин, 1984, с. 79]. В то же время, как заметил Т. Кун, "то, что кажется регрессивным сдвигом проблемы, может быть началом весьма длительного периода прогресса" [Цит. по: Лакатос, 1978, с. 228].

Схема Лакатоса тоже представляет собой удобную модель для самоописания науки. Предложенная в 1970 г., через несколько лет после концепции Куна, методология научных исследовательских программ быстро завоевала столь же широкую популярность. В обзорной и исторической литературе по искусственному интеллекту часто встречается терминология "парадигм" или "исследовательских программ". Насколько верно используется она при этом, в том ли смысле, который вкладывали в эти понятия Лакатос и Кун? Каковы результаты применения данных моделей, дают ли они удовлетворительное объяснение динамики исследований? Ответы на эти вопросы мы попробуем дать ниже, рассмотрев конкретные примеры использования этих логико-методологических схем в работах как самих специалистов по искусственному интеллекту, дающих рефлексивную картину исследований "изнутри", так и сторонних наблюдателей, избравших данную область в качестве предмета историко-научного анализа.

### **1. 2. 2. Парадигмы в истории искусственного интеллекта**

Термин "парадигма" используется в исторических и обзорных описаниях области искусственного интеллекта чрезвычайно часто, возможно, даже слишком. Во всяком случае, канадский исследователь Л. Шуберт в ответ на вопрос анкеты, что бы он предпринял, если бы был назначен "философским королем" научного сообщества искусственного интеллекта (опрос проводился в 1978 г. ), ответил, что запретил бы употреблять слово "парадигма", "которое всякий полощет, как ему вздумается, как только отыщет его в словаре Уэбстера" [Special..., 1980, p. 110].

О парадигмах в истории искусственного интеллекта говорят в основном лишь в тех случаях, когда описывают отдельные рубежные моменты, характеризуя их как смену парадигм. Состоянию "нормальной науки" уделяется гораздо меньше внимания. Были,

однако, и попытки представить историю искусственного интеллекта в целом как последовательность сменяющихся парадигм. В 1984 г. английский исследователь Р. Форсайт изобразил сжатую историю данной области в форме таблицы [Форсайт, 1987, с. 10]:

Годы	Парадигма	Исполнители (теоретики)	Система
50-е	Нейронные сети	Розенблатт (Винер, Маккалок)	PERCEPTRON
60-е	Эвристический поиск	Ньюэлл и Саймон (Шеннон, Тьюринг)	GPS
70-е	Представление знаний	Шортлифф (Минский, Маккарти)	MYCIN
80-е	Обучающиеся машины	Ленат (Сэмюэль, Холланд)	EURISCO

Со схемой Форсайта согласуется введенное П. Уинстоном различие парадигмы "Универсального решателя задач" (GPS) и парадигмы систем продукций (MYCIN) [Уинстон, 1980, с. 15]. На начало 70-х гг. как на момент смены парадигм указывают и отечественные авторы А. Л. Горелик и И. Е. Гуревич: "способность 'интеллектуальной' системы решать задачи определяется в первую очередь теми специальными знаниями, которыми она располагает, и только во вторую – универсальностью и мощностью используемых ею методов вывода. Парадигма искусственного интеллекта сменилась в том отношении, что фундаментальной проблемой стало считаться не обнаружение ряда мощных методов, а способ представления значительного объема знаний таким образом, чтобы были обеспечены эффективный доступ к ним и эффективное их использование" [Горелик и др., 1982, с. 17].

Дж. Шопмен, как и Форсайт, ассоциировал парадигмы в истории искусственного интеллекта с определенными периодами [Question..., 1988]. Однако, согласно мнению "изнутри" области, принадлежащему Дж. Маккарти, в истории данной дисциплины не было периодов полного доминирования одной парадигмы; делались лишь более сильные упоры, акценты на некоторых направлениях, но разнообразие точек зрения всегда сохранялось. Маккарти даже предложил выделить в исследованиях по искусственному интеллекту "субпарадигмы" [Annals..., 1988, p. 228].

Статуса "парадигмы" удостоились у С. Адданки активно развивающиеся в последние годы исследования по параллельным взаимодействующим сетям, известные под обобщающим наименованием "коннекционизм" [Addanki, 1987, p. 200].

Наступающую смену парадигм сразу в трех подобластях искусственного интеллекта – логическом рассуждении, представлении знаний и системах общения – отметил Д. А. Поспелов: "Основная логическая парадигма текущего момента в интеллектуальных системах – извлечение программы решения задачи из некоторого доказательства. Она предполагает, что интересующая нас задача может быть представлена в виде задачи доказательства теоремы. ... Новая парадигма заменяет логический вывод обоснованием в имеющейся системе знаний... На следующем шаге возможна иная парадигма, когда будет использоваться оправдание решения в имеющейся системе ценностей" [Поспелов Д. А., 1988, с. 29-31]. В исследованиях по представлению знаний, пишет Д. А. Поспелов, происходит переход от символьных моделей к визуальным. В системах общения парадигма понимания отдельных предложений сменяется лингвистикой целостного текста.

В другой подобласти искусственного интеллекта – машинном обучении – были также выделены парадигмы, сменяющие друг друга на различных этапах развития исследований [Michalski et al., 1984, p. 14-16]:

1950-е годы – нейронное моделирование (работы Ф. Розенблатта и других) и подход теоретического решения (А. Сэмюэль);

1960-е – обучение, ориентированное на символьные понятия (Э. Хант, П. Уинстон);

1970-е – самообучение с генерацией обучающих заданий (Д. Ленат).

Заметим, что работы Д. Лената и А. Сэмюэля, отнесенные Р. Форсайтом к одной парадигме, в данной схеме оказались в разных. Кроме того, как мы видим, в 1950-е годы конкурировало два альтернативных подхода, не образующих единую парадигму.

Можно привести и другие примеры, но уже видны основные недостатки применения концепции парадигм к истории искусственного интеллекта. Сразу бросаются в глаза чрезвычайно малые отрезки времени (порядка 10 лет), отделяющие одну "научную революцию" от другой. Согласно Т. Куну, различные парадигмы – это несоизмеримые способы видения мира [Кун, 1975, с. 20], неявно определяющие круг проблем и методов исследования для последующих поколений ученых, на долгое время отвращающие их от конкурирующих моделей научных исследований [там же, с. 27]. Вряд ли даже самое динамичное и свободное от консерватизма научное сообщество способно было бы трижды сменить "способ видения мира" на памяти одного лишь поколения ученых!

Перечисленные выше концептуальные сдвиги сыграли огромную роль в истории искусственного интеллекта, но все же не могут претендовать на статус "научных революций". Скорее это похоже на смену "защитных поясов" исследовательской программы в концепции И. Лакатоса. Возьмем, к примеру, цепочку новаций, выстроенную Д. А. Поспеловым: "Логический вывод в базе данных – Обоснование в системе знаний – Оправдание в системе ценностей". Здесь сохраняется "жесткое ядро" (формулировка интеллектуальной проблемы в виде доказательства некоторого утверждения), а развитие происходит согласно "позитивной эвристике": использовать менее формальные способы рассуждения, в большей мере учитывать неявные и интуитивные аспекты мышления.

Таким образом, примеры использования схемы Т. Куна в объяснении развития исследований по искусственному интеллекту не выглядят убедительными.

Быть может, при более аккуратном использовании эта модель дала бы результат? По нашему мнению, к области искусственного интеллекта концепцию парадигм применять еще рано. Братья Дрейфусы уже в 1980-е годы констатировали, что исследователи зачастую не знают наверняка, что может служить темой диссертации или приемлемой

публикации в данной области, поскольку до сих пор нет согласия в том, каковы же насущные проблемы [Dreyfus et al., 1986, p. 99]. Это означает, что в области не сложилась система "головоломок", образцы решения которых хорошо известны. По результатам опроса специалистов по представлению знаний, лишь 7% из них используют уже существующие схемы представления, остальные же разрабатывают свои собственные, оригинальные [Special..., 1980, p. 50]. Как заметил Т. Кун, "для допарадигмального периода в особенности характерны частые и серьезные споры о правомерности методов, проблем и стандартных решений..." [Кун, 1975, с. 72]. Именно такова ситуация в области искусственного интеллекта. Поэтому, на наш взгляд, исследования в данной области еще не вступили в пору "нормальной науки", они находятся лишь на "допарадигмальной" стадии.

Более фундаментальную причину неприменимости логико-методологической схемы Куна к истории информационной науки (центральной частью которой является искусственный интеллект) указывает А. И. Ракитов. В данной схеме, считает он, "глубокий синтез столь различных наук, как физика, лингвистика, психология, эпистемология, математика, технические науки и т. д., связанных в контексте информационной технологии, не мог получить адекватного методологического осмысления. Короткая, но очень насыщенная история информационной науки показывает, что эта новая наука не укладывается в понятие парадигмы, а ее развитие не носит парадигмального характера... Синтез какой-либо парадигмы из теорий, концепций, установок и методов другой науки невозможен в силу некумулятивности различных парадигм как в синхронном, так и в диахронном планах... компоненты новой информационной науки были парадигмально зрелыми еще до начала компьютерной революции, до становления информационной технологии и возникновения информационной науки" [Ракитов, 1987, с. 39].

Столкновение на поле исследований по искусственному интеллекту различных дисциплинарных парадигм, действительно, сильно осложняет ситуацию, порождает слабоструктурированное разнообразие терминологии, концептуальных схем, методов, образцов работы. Следует, однако, учесть, что при этом взаимодействуют не сами исходные парадигмальные структуры из различных дисциплин, а их "отпечатки", "слепки", функционирующие уже в контексте единой дисциплины. Возникает основа для установления определенных соглашений, согласования терминов, сравнения методов. Этого еще не достаточно для образования единой парадигмы, но уже появляется базис для сравнения и конкуренции различных подходов. Последние, вероятно, могут выступать в данном случае в качестве самостоятельных исследовательских программ.

### **1. 2. 3. Научные исследовательские программы в истории искусственного интеллекта**

Наиболее последовательно методологию исследовательских программ при анализе истории искусственного интеллекта проводят в своих работах братья Дрейфусы – философ Х. Дрейфус и специалист по информатике С. Дрейфус. В самом крупном масштабе эта история представляется им борьбой двух исследовательских программ – коннекционизма и информационной (символьной) обработки. В рамках коннекционизма разрабатываются нейроноподобные модели параллельного действия, предназначенные для обработки целостных образов. Символьный подход ("обычный искусственный

интеллект") предполагает последовательную переработку информации, при которой каждому отдельному свойству образа отвечает определенный элемент системы.

Симпатии Дрейфусов — целиком на стороне коннекционизма. Они пишут: "обычный искусственный интеллект как информационная обработка выглядит прекрасным примером того, что Имре Лакатос назвал бы регрессирующей исследовательской программой. Она началась достаточно благоприятно с ранней работы Ньюэлла и Саймона, показавшей, что можно запрограммировать компьютер так, чтобы смоделировать определенные формы человеческого манипулирования символами. С успехами в начале 70-х она быстро установилась как преуспевающая исследовательская программа. Затем, довольно неожиданно искусственный интеллект столкнулся с трудностями... Сомнения возникли в связи с провалом попыток запрограммировать понимание детских рассказов, и вскоре стало очевидно, что реализация здравого смысла — серьезная проблема" [Dreyfus et al., 1986, p. 98].

Следует заметить, что коннекционистская программа и близко не подобралась к проблемам такой сложности, как понимание естественного языка, так что не вполне понятно, по какой шкале Дрейфусы отмеряют успехи и провалы. Вывод, тем не менее, они делают из этих "неудач" глобальный: "Благодаря исследованиям по искусственному интеллекту рассуждения Платона и Канта о том, что разум действует по правилам, нашли, наконец, эмпирическую проверку в попытке использовать логические машины для осуществления человеческого понимания. После двухтысячелетнего усовершенствования традиционный взгляд на разум оказался неадекватным" [Ibidem].

Относительность оценок, даваемых Дрейфусами, иллюстрируется элементарным сопоставлением с их же высказываниями более раннего периода. Сейчас начало 70-х представляется им временем "преуспевания" исследований, но ведь именно в 1972 г. Х. Дрейфус выпустил книгу "Чего не могут вычислительные машины?" [Дрейфус, 1978], в которой "разочарование", "крушение иллюзий" и "пессимизм" были еще самыми мягкими из характеристик области искусственного интеллекта.

Разберем хотя бы один пункт критики Дрейфусом работ А. Ньюэлла и Г. Саймона. Эти исследователи описали одну из трудностей, возникающих при работе их системы "Универсальный решатель задач" (GPS), — расхождение последовательности действий машины с протоколом решения той же задачи человеком: "Здесь GPS просто применила правило к текущему выражению, в то время как испытуемый вернулся назад и исправил предыдущее применение. В программе мы не найдем чего-либо, что соответствовало бы этому действию" [Ньюэлл и др., 1967в, с. 298-299]. Дрейфус сразу видит в этом эпизоде регресс исследовательской программы: "Ньюэлл и Саймон, по-видимому, не осознают, что этот отход от схемы действий машины ... чреват для их теории таким же крушением, которое потерпела теория Птолемея, когда обнаружилось, что действительные траектории планет ей не соответствуют. ... подобно древним астрономам, они пытаются спасти свою теорию, добавляя к ней пару-другую эпициклов. Они по-прежнему полагают ... что в основе подобного механизма лежат просто более изощренные методы поиска и что если дать GPS возможность 'постоянно немного оглядываться назад, на свои предыдущие действия', то все будет в порядке" [Дрейфус, 1978, с. 60]. "Здесь мы также сталкиваемся с чем-то вроде залатывания дыр и подшивания кусочков с помощью приемов ad hoc... это своего рода карабканье на дерево в надежде добраться до Луны" [там же, с. 66].

Заметим, что система Птолемея отнюдь не терпела крушения, пока накапливала эпициклы, — наоборот, она доказывала свою хорошую восприимчивость к аномалиям.

Крах ждал ее в столкновении с конкурирующей теорией. Но если бы вдруг случилось так, что гелиоцентрическая система никогда бы не возникла, то вряд ли у нас были бы основания говорить о теории Птолемея с подобным пренебрежением. Тем более непонятен критицизм Дрейфуса по отношению к работам Ньюэлла и Саймона. Разве он может указать систему, моделирующую поведение человека в данной ситуации лучше, чем "Универсальный решатель задач"?

На наш взгляд, применение Дрейфусами методологии научных исследовательских программ к истории искусственного интеллекта было проведено недостаточно корректно, вследствие чего оценки отдельных исследований оказались субъективными, а общая картина – искаженной. По Лакатосу, оценить прогресс или вырождение исследовательской программы можно лишь по "внешнему" критерию – в сравнении с другой, конкурирующей программой. Успех одной программы есть неудача другой, и наоборот. Дрейфусы же пытаются использовать "внутренний" критерий — сравнение с поведением человека. Но мы же не можем, например, оценить адекватность физической теории лишь по абсолютной величине точности ее предсказаний, вне сравнения с точностью предсказания других теорий. Так же и в области искусственного интеллекта – нельзя сказать, хорошо или плохо данная система моделирует поведение человека, если нет конкурирующей модели.

Разделяя с Дрейфусами принципиальное убеждение в применимости методологии исследовательских программ к истории искусственного интеллекта, мы считаем, однако, что это применение должно базироваться на более тщательном учете особенностей области и требует соответственной настройки методологического инструментария.

Подводя итог осуществившимся попыткам применить к истории искусственного интеллекта различные логико-методологические схемы, сошлемся на мнение А. Ньюэлла: "Использование парадигм Куна и исследовательских программ Лакатоса мало что дает. Можно показать, что на протяжении своей короткой истории искусственный интеллект развил одну, максимум две парадигмы. Аналогично, он содержит лишь горстку исследовательских программ. Для анализа, однако, необходимо достаточно большое число примеров. Во внутренней истории искусственного интеллекта, несомненно, слишком мало парадигм. Вероятно, это верно и для исследовательских программ, хотя было бы интересно попытаться дать такое описание искусственного интеллекта" [Newell, 1983, p. 188].

По нашему мнению, даже небольшое число исследовательских программ способно за счет активной конкуренции породить сложную и богатую событиями картину развития исследовательской области. Нам, как и А. Ньюэллу, тоже интересно взглянуть на эту картину.

### **1. 3. Характеристика адекватной модели развития области**

#### **1. 3. 1. Сходство со схемой И. Лакатоса**

Как уже отмечалось выше, схема развития науки по Лакатосу предусматривает постоянную конкуренцию научных исследовательских программ, каждая из которых развивается путем смены вспомогательных теорий ("защитных поясов"), выдвигаемых "позитивной эвристикой", при неизменном "жестком ядре". В исследованиях по искусственному интеллекту легко обнаружить прямые аналоги этих компонент схемы.



В основе любой системы искусственного интеллекта лежит некоторая исходная модель мышления, включающая описание определенного класса интеллектуальных задач и механизм поиска их решения. Различные исходные модели в самой общей, качественной форме черпаются из смежных с искусственным интеллектом областей – психологии, лингвистики, логики, нейрофизиологии, биологии. Дальнейшая работа специалистов по искусственному интеллекту состоит в разработке адекватных формальных описаний и конкретных механизмов реализации этих моделей. Каждая модель становится "ядром" соответствующей исследовательской программы.

Подробнее содержание исходных моделей будет рассмотрено в следующей главе. Здесь же мы укажем лишь на некоторые проблемы, связанные с выделением "ядра" исследований по искусственному интеллекту. Дело в том, что такая попытка, не вполне, на наш взгляд, удачная, уже была предпринята Х. Дрейфусом [Дрейфус, 1978, с. 106]. Он эксплицировал четыре допущения, лежащие в основе искусственного интеллекта:

1. Биологическое допущение: переработка информации на нейронном уровне носит дискретный характер по типу электрических переключательных схем.

2. Психологическое допущение: мышление есть переработка информации, заданной в бинарном коде, по формальным правилам.

3. Эпистемологическое допущение: все знания могут быть формализованы и выражены в терминах булевых функций.

4. Онтологическое допущение: все происходящее в мире можно представить в виде множества фактов, каждый из которых логически не зависит от остальных.

Дрейфус, как видно, считает, что эти положения, явно или неявно, принимаются любым специалистом по искусственному интеллекту. Можно критиковать спорность самих формулировок (скажем, информация не обязательно представляется в бинарном коде, а онтологическое допущение никак не согласуется с активным использованием в современных исследованиях неклассических (модальных, каузальных, псевдофизических) логик), но сейчас для нас важен сам факт объединения Дрейфусом всех этих допущений в единый блок.

На наш взгляд, биологические, психологические и логические модели в искусственном интеллекте никогда не уживались вместе. Подобный синкретизм свойствен ранним работам У. Маккалока и У. Питтса [Маккалок и др., 1956] – кибернетическому подходу, послужившему для исследований по искусственному интеллекту лишь точкой отталкивания. Своеобразие искусственного интеллекта в сравнении с кибернетикой как раз и состояло в отказе от попыток совместить нейрофизиологическую и феноменологическую картины мышления и переходе целиком на феноменологический уровень. Что же касается психологического и эпистемологического (логического) подходов, то несовместимость психологических моделей человеческой субъективности и эпистемологических моделей объективного знания также была ясна специалистам по искусственному интеллекту с самого начала.

Соединение несовместимого понадобилось Дрейфусу, по-видимому, из-за того, что он считал "обычный искусственный интеллект" единой исследовательской программой символьной переработки. Как мы покажем ниже, в этой области действует несколько исследовательских программ с принципиально различными "ядрами". Именно отличия "ядер" позволяют ввести критерий разделения программ, т.к. многие частные схемы могут использоваться в равной мере представителями разных программ.

Роль "защитных поясов" играют практические реализации исходных качественных

моделей – конкретные системы искусственного интеллекта, компьютерные программы, выполняющие определенные интеллектуальные функции. По сути дела, они выступают в качестве частных теорий мышления в рамках исходных постулатов, заложенных в "ядро". Действительно, как указывает Дж. Вейценбаум, "теории – это тексты. Тексты записываются на некотором языке. Машинные языки – это тоже языки, и теории можно на них записывать... Тексты, записанные в виде программ вычислительной машины, – это обычные теории" [Вейценбаум, 1982, с. 195-196]. А. Ньюэлл и Г. Саймон рассматривали "Универсальный решатель задач" именно как теорию: "Предлагается теория, объясняющая, каким образом некоторые люди пытаются решать простые формальные задачи. Исследования, послужившие основой для данной теории, тесно связаны с проблемами обработки информации и построения "разумных" автоматов, а сама теория выражена в виде программы для вычислительной машины" [Ньюэлл и др., 1967в, с. 283]. Авторы задают и критерий оценки таких систем-теорий: "Если поведение программы будет достаточно близким к зафиксированному в протоколе поведению испытуемого, такая программа даст нам хорошую теорию процесса решения задач человеком" [там же, с. 289]. Разработка более совершенных с данной точки зрения систем искусственного интеллекта, опирающихся на единую исходную модель, соответствует смене теорий в рамках одной исследовательской программы при сохранении общего "жесткого ядра".

"Позитивная эвристика", которая, по Лакатосу, определяет проблемы для исследования и выделяет "защитный пояс", в данном случае состоит из двух компонент. Первая связана с областью, из которой почерпнута исходная модель, и задает перечень интеллектуальных проблем, характерных для этой области. Скажем, лингвистическая модель мышления влечет за собой выбор характерных задач — понимание естественного языка, машинный перевод, а психологическая модель диктует использование обычных ситуаций психологических экспериментов на запоминание или решение логических головоломок. Вторая компонента связана с областью информатики, в рамках которой развиваются исследования по искусственному интеллекту. Она определяет выбор средств реализации конкретных интеллектуальных систем, учитывая аппаратные возможности ЭВМ и ее программное обеспечение (включая машинные языки). Имеется и обратная связь: задачи искусственного интеллекта вызывают к жизни новые языки программирования (например, LISP и PROLOG) и даже новые аппаратные решения (т. наз. ЛИСИП-машина).

Сходство схемы развития области искусственного интеллекта с моделью Лакатоса очевидно, но специфика этой области требует особой "настройки" модели, конкретизации ее общих положений для данного случая.

### **1. 3. 2. Отличительные особенности модели**

Мы остановимся на трех основных особенностях исследовательских программ в области искусственного интеллекта. Первая связана с характером самих исследований, их научно-инженерной двойственностью и, соответственно, "обратимостью" их продукта, который может выступать и как инженерный проект, и как научное знание. Вторая характерная особенность определяется предметом исследований – интеллектуальным "пространством", образованным "решеткой" интеллектуальных задач и методов моделирования. Разнообразие возможных траекторий движения по этому пространству заставляет иначе взглянуть на проблему конкуренции программ. Со спецификой этой

конкуренции, определением критериев прогресса и регресса исследовательских программ как раз связана третья рассматриваемая особенность.

Как уже отмечалось выше, научное сообщество столкнулось с проблемой двойственности исследований практически с самого начала. Первой реакцией на эту трудность явились попытки ввести различие "искусственного интеллекта" (как деятельности по созданию "разумных" машин) и "моделирования процессов познания" (как создания теории человеческого мышления с помощью его моделирования на ЭВМ). Первый сборник статей по искусственному интеллекту "Вычислительные машины и мышление", вышедший в 1963 г., был поделен его редакторами Э. Фейгенбаумом и Дж. Фельдманом на две соответствующие части [Вычислительные..., 1967].

Сразу, однако, возникла проблема – к какой из частей отнести цикл работ А. Ньюэлла, Г. Саймона и Дж. Шоу по созданию систем, играющих в шахматы, доказывающих теоремы логики и решающих другие задачи, но одновременно и моделирующих сам процесс человеческого мышления. Редакторы сборника так охарактеризовали одну из работ этого цикла – шахматную компьютерную программу: "Программа Ньюэлла, Шоу и Саймона относится к тем исследованиям, которые лежат в пограничной области между проблемой искусственного интеллекта и моделированием процесса решения задач человеком. Строго говоря, она вовсе не задумана как некая модель решения задач человеком в шахматной ситуации. Однако, разрабатывая свою программу, авторы исходили из убеждения, что человеческие методы решения задач, характеризующиеся высокой степенью адаптивности и избирательности, должны оказаться более эффективными при решении шахматных задач, чем другие предложенные вычислительные схемы. Работа их шахматной программы как будто подтверждает правильность этого убеждения" [там же, с. 32-33].

Для того, чтобы инженерная разработка была эффективной, оказалось необходимым моделировать некоторые аспекты человеческого мышления, что уже представляло собой научную задачу. Научная и инженерная цели вступили в тесную взаимосвязь.

В конечном счете редакторы отнесли шахматную программу и систему "Логик-теоретик" [Ньюэлл и др., 1967а, 1967б] к первой части сборника, а "Универсальный решатель задач" [Ньюэлл и др., 1967в] – ко второй, хотя все три системы базировались на едином принципе эвристического поиска. Фактически одна и та же система искусственного интеллекта, рассматриваемая с разных точек зрения, выглядела решением то инженерной задачи, то научной проблемы.

Позднее работы из обеих частей этого сборника стали равно рассматриваться как классические системы искусственного интеллекта. И сам Э. Фейгенбаум вскоре уже включил "моделирование процессов познания" в область искусственного интеллекта [Фейгенбаум, 1973]. Для многих ведущих специалистов стала ясна искусственность прежнего размежевания. Крупнейший исследователь, участник Дартмутского проекта, профессор Массачусетского технологического института Марвин Минский писал в 1974 г.: "В настоящем очерке я не провожу границы между теорией мышления человека и схемой построения интеллектуальной машины. Их разделение сегодня не пошло бы никому на пользу" [Минский, 1979, с. 5]. В 1978 г., при ответе на вопрос анкеты "предназначена ли Ваша система быть инструментом для построения интеллектуальных систем или же самодостаточным исследованием теоретических проблем?", другие видные американские ученые Д. Бобров и Т. Виноград констатировали неразделимость этих функций [Special..., 1980, p. 52].

Наша точка зрения состоит в том, что особенностью исследовательских программ в области искусственного интеллекта является их "двуликость" в отношении инженерных или научных целей исследователей, работающих рамках данной программы. Модель мышления, заложенная в "жесткое ядро", может выступать и как основа теории мышления, и как база инженерного проекта. Так же двойственен и конечный продукт – интеллектуальная система, которая потенциально является и записанной на машинном языке теорией, и технической реализацией проекта. Американский специалист в области когнитивной психологии Дж. Миллер еще в 1962 г. указал на двоякую функцию компьютерных программ искусственного интеллекта – как "модели интеллектуальной системы и, даже чаще, вида языка, на котором можно выразить теорию" [Цит. по: Newell, 1970, р. 366]. Двойственность исследований в данной области подчеркнул и Б. В. Бирюков, выделивший две главные цели этих работ: "Одна из них – моделирование на цифровой машине некоторых существенных черт интеллектуального (разумного) поведения. Другая – использование ЭВМ для решения сложных проблем, овладение которыми до сих пор было под силу только человеку. Различие этих двух целей относительно (многие кибернетические разработки подчинены обоим целям одновременно) и отражает лишь разные акценты" [Бирюков, 1978, с. 301]. Заметим, что, хотя различие этих целей может подчас выглядеть и абсолютным (одни авторы открыто декларируют свою приверженность только первой цели, другие же – только второй), сам результат исследований можно реально рассматривать как воплощение любой из указанных целей.

Мы сталкиваемся здесь с явлением, названным М. А. Розовым "рефлексивной симметрией знания". Ближе всего к нашему случаю пример симметрии предметного и методического знания. Розов характеризует его так: "Представьте себе описание конкретного эксперимента, например, эксперимента по определению скорости света. К этому описанию можно относиться двумя различными способами. Все зависит от того, интересуется ли нас свет и скорость его распространения или метод измерения этой скорости. В первом случае описание будет выглядеть как знание о свете, во втором – как описание образца деятельности, которую можно по этому описанию воспроизводить. Одно и то же описание воспроизводится как знание двух разных типов. Это либо предметно ориентированное знание, представляющее собой характеристику какого-либо явления, либо знание-рецепт, знание-метод" [Розов, 1986, с. 14]. Выбор того или иного описания определяется рефлексией ученого, поэтому симметрия и является "рефлексивной". Аналогично, в исследованиях по искусственному интеллекту мы обнаруживаем симметрию предметного знания (о механизмах мышления) и технического знания – метода инженерного проектирования (как построить интеллектуальную машину).

Данная ситуация мгновенно порождает проблему оценки конкретных исследований. Действительно, некоторая интеллектуальная система может хорошо моделировать процесс человеческого мышления, но при этом быть чрезвычайно неэффективной с инженерной точки зрения, т. е. просто плохо решать задачи (ведь человек не всегда мыслит эффективно). А хороший инженерный проект, в свою очередь, может неадекватно воспроизводить человеческий метод мышления. Автор конкретной системы искусственного интеллекта не испытывает таких колебаний, он выбирает для себя одну – научную или инженерную – цель и оценивает работу системы по однозначному критерию. Но с методологической точки зрения такое решение выглядит субъективным. Чтобы придать критериям оценки объективность, следует обратиться к общей картине динамики

исследовательских программ в рамках всей области.

Выше мы уже описывали структуру того интеллектуального "пространства", в котором прокладывают себе дорогу исследования по искусственному интеллекту. Оно имеет форму "решетки" или таблицы, по столбцам которой расположены интеллектуальные задачи, а по строкам — методы моделирования мыслительных функций, позволяющие искать решение этих проблем. С развитием исследований клетки этой таблицы постепенно заполняются — появляются системы, в которых соответствующий метод применяется к соответствующей задаче. Аналогичную картину описания области предложил А. Ньюэлл: "История искусственного интеллекта в целом может быть написана в терминах географии задач, успешно выполненных системами искусственного интеллекта. Почти ортогонально к этому направлению задач лежит направление интеллектуальных функций, необходимых для интеллектуальных систем — репрезентация, методы решения проблем, распознавание, получение знаний и т.д. — то, что можно определить как физиологию интеллектуальных систем. Так, история может быть написана в форме возрастающего понимания и механизации каждой функции. Оба этих структурных качества искусственного интеллекта и в особенности, возможно, их матрица обеспечивают достаточно плодотворный базис для истории" [Newell, 1983, p. 189]. Сходным образом структурируется, например, и одна из подобластей искусственного интеллекта — машинное обучение. Исследования при этом классифицируются по трем критериям — по лежащей в их основе стратегии обучения, по способу представления знаний или навыков и по областям применения. "Каждая точка в этом трехмерном пространстве означает определенную обучающую стратегию, использующую определенное представление знаний и примененную к определенной области" [Michalski et al., 1984, p. 7]. Если объединить стратегию обучения и способ представления знаний в едином понятии метода моделирования, то мы вновь получим структуру двумерного пространства.

Важно отметить, что в данном пространстве можно двигаться двояким образом — по методам или по задачам. Любая система искусственного интеллекта как бы потенциально включена в эти две традиции — поиска лучшего метода решения данной задачи и исследования возможностей определенного метода на множестве задач. Здесь вновь, но уже на другом уровне, возникает предметно-методическая симметрия знания.

Две указанные традиции четко различили Дж. Маккарти и П. Хейес: "Были написаны программы для некоторого класса задач, требующих для своего решения сложной умственной деятельности человека. В ходе разработки этих программ были установлены более или менее общие механизмы интеллекта... Нередко испытание программ приводило к лучшему пониманию интеллектуальных механизмов и к выявлению новых механизмов... Альтернативный подход состоит в том, чтобы начинать с механизмов интеллекта (например, механизмов памяти, принятия решения на основе взвешенных сумм субкритериев, обучения, разветвленного поиска, экстраполяции), а затем разрабатывать задачи, к которым можно применить эти механизмы" [Маккарти и др., 1972, с. 41-42]. Каждый отдельный ученый, как правило, ограничивается движением лишь в одном направлении, и тогда его рефлексия может расходиться с рефлексией коллеги, движущегося "перпендикулярно". Как замечают те же авторы, исследователи не охватывали проблемы в целом и думали, что видят лес, когда на самом деле смотрели лишь на одно дерево" [там же, с. 42].

Хотя данное интеллектуальное пространство едино и должно служить полем

столкновения всех исследовательских "программ, каждая программа имеет в нем свои "предпочтительные" траектории. Они проходят по задачам, связанным с той областью, из которой заимствована исходная модель мышления, составляющая "жесткое ядро" данной программы. Скажем, логистическая программа в большей мере ориентирована на доказательство теорем, чем на понимание естественного языка, а программа, основанная на лингвистической модели, ведет себя ровно наоборот. Конечно, исследовательские программы вторгаются и на "чужие" территории, иначе соревнование их было бы вовсе невозможно. Но эта конкуренция похожа не на гонки автомашин по одной трассе, а скорее на блуждание грибников в одном лесу.

Неудивительно поэтому, что проблема выбора критериев сравнения различных систем искусственного интеллекта стоит в научном сообществе очень остро. Из-за различия исходных позиций (каждый, так сказать, входит в лес со своего конца) никак не удастся выработать единую точку зрения. По данным опроса 1978 г., 60% специалистов по искусственному интеллекту считают, что критерии, по которым следует оценивать их собственные системы, не являются универсально применимыми, причем треть из них рассматривает эти критерии как совершенно уникальные, годящиеся лишь для одной системы [Special..., 1980, p. 66]. Специалисты из компании "Ксерокс" К. ван Лейн, Дж. С. Браун и Дж. Грино справедливо отметили "отсутствие 'соревновательной аргументации', с помощью которой результаты одного моделирования можно было бы сравнить с результатами другого. Без такой аргументации теории искусственного интеллекта стоят на носках, а не на плечах своих предшественников" [Цит. по: Gardner, 1985, p. 178].

В этом случае, по нашему мнению, критерием оценки теорий (интеллектуальных систем) в рамках одной исследовательской программы может выступать величина покрываемого ими интеллектуального пространства (реализуемых методов и решаемых задач). Одна теория, как у И. Лакатоса, сменяет другую, если обладает по отношению к ней дополнительным содержанием [См.: Лакатос, 1978, с. 221]. Критерий же в борьбе разных исследовательских программ – новизна выдвигаемых задач и методов. Прогрессировать будет та исследовательская программа, которая расширяет само интеллектуальное пространство, а не замыкается в уже известном объеме. Из этого следует, что в определенные периоды могут прогрессировать несколько программ, расширяя пространство в разные стороны и не вступая до поры в конкуренцию.

Борьба исследовательских программ разворачивается одновременно в двух плоскостях — научной и инженерной. При этом одна программа может с точки зрения научных достижений прогрессировать, а по инженерным критериям – вырождаться, а другая – наоборот. Все эти парадоксы воочию продемонстрирует реальная история искусственного интеллекта, к рассмотрению которой мы приступаем.

## Глава 2. ИСТОРИЧЕСКАЯ ДИНАМИКА ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

### 2. 1. Истоки и зарождение исследовательских программ

#### 2. 1. 1. Кибернетический контекст

Фундаментальный методологический сдвиг в середине XX в., изменивший облик многих научных дисциплин и, в частности, создавший условия для возникновения области искусственного интеллекта, был связан с появлением кибернетики. Выход в 1948 г. первого издания "Кибернетики" Н. Винера ознаменовал переход на новые основания для объяснения живой природы. Был предложен механизм осуществления целесообразной деятельности – отрицательная обратная связь. От энергии (центрального понятия ньютоновской механики) внимание переместилось к информации. Основные понятия информационной теории (кодирование, память, шум) объясняли широкий спектр явлений — от функционирования электрических цепей до поведения разумных существ. Смысл кибернетического объяснения состоял в редукции сложного поведения человека и животных, функционирования их органов к действию простых механизмов обратной связи, уже достаточно изученных и давно применявшихся в технических устройствах. Важный методологический шаг был сделан в результате перехода от закрытых ньютоновских систем к открытым, обменивающимся энергией и информацией со средой.

Кибернетические идеи столь стремительно вошли в научный обиход, что в предисловии ко второму изданию "Кибернетики" 1961 года Винер даже выразил опасение, что ныне "книга может показаться банальной. Значение обратной связи в техническом проектировании и в биологии твердо установлено. Значение информации и методика ее измерения и передачи составляют целый предмет изучения для инженера, физиолога, психолога и социолога" [Винер, 1968, с. 29]. В фундамент искусственного интеллекта был заложен первый камень: установлено, что люди и машины – лишь разновидности устройств по переработке информации.

Следующий шаг состоял в создании машины, которая по характеру и объему такой переработки была бы сопоставима с человеком. Одним из первых параллель "человек – компьютер" установил Дж. фон Нейман, в 1940-50-е годы создававший в Принстонском институте высших исследований проекты первых ЭВМ. Уже в своих ранних работах он сравнивает элементы человеческой нервной системы с частями компьютера: память, управляющие органы, чувствительные (афферентные) и действующие (эфферентные) нейроны (в качестве входного и выходного устройств ЭВМ). В лекции 1948 г. фон Нейман подробно разбирает сходства и отличия вычислительных машин и живых организмов [фон Нейман, 1960б]. Будучи автором самой идеи антропоморфизации компьютеров, он в то же время предостерегал от преувеличения ее значимости. Основные возражения сводились к следующему:

- компьютер, в отличие от человека, состоит из больших и ненадежных элементов;
- ЭВМ потребляет существенно больше энергии, чем человек;
- живые организмы демонстрируют как дискретные (взаимодействие нейронов), так и непрерывные (ток крови) механизмы, в то время как компьютер может быть только либо целиком дискретным (цифровым), либо целиком непрерывным (аналоговым).

С содержательной точки зрения эти конкретные возражения сейчас представляются

во многом наивными (ничто в мире не совершенствуется и не миниатюризуется с такой скоростью, как компьютеры), но для нас важна методологическая посылка фон Неймана – сравнение человека и компьютера лишь по "аппаратурным" критериям, на уровне физиологии и "железок". На этом уровне человек и машина были и всегда будут принципиально различны, просто как явления живой и неживой природы. Бессмысленно стремиться достичь "объема памяти" или "быстродействия" человека, ибо неясно, как же из нейронных взаимодействий, изучаемых нейрофизиологией, вдруг рождается мысль, которую и описать-то в нейрофизиологических понятиях невозможно. До тех пор, пока этот переход не прояснен, можно сколь угодно точно моделировать функционирование человеческого мозга, но так и не получить мышление.

Бесперспективность "аппаратурного" человеко-машинного параллелизма хорошо понимал английский кибернетик А. Тьюринг. Подобный путь, по его мнению, аналогичен затратам усилий на создание машины, ходящей на ногах, вместо использования колес. Тьюринг предложил "провести четкое разграничение между физическими и умственными возможностями человека" [Тьюринг, 1960, с. 20]. В статье, впервые опубликованной в 1950 г., он предложил процедуру, вошедшую позднее в научный обиход под названием "тест Тьюринга", для оценки "разумности" поведения вычислительной машины. Человеко-машинный параллелизм при этом был перенесен с аппаратурного на феноменологический уровень; сравнивались не структуры, а функции. В тесте Тьюринга от машины требовалось лишь умение выдавать себя за человека в разговоре с абонентом по телетайпу [там же, с. 19-20]. Каким образом ей удастся этого достичь, из каких "железок" она собрана – все это в данном случае совершенно не важно. Идея Тьюринга о функциональном моделировании человеческого разумного поведения на компьютере стала последним блоком в фундаменте искусственного интеллекта. Точнее говоря, могла бы стать. В жизни все оказалось несколько сложнее.

Дело в том, что в начале 50-х годов идеи Тьюринга в США (да и в самой Великобритании) не были широко известны. Как утверждает Дж. Маккарти, никто из "пионеров искусственного интеллекта" в те годы эту статью Тьюринга не читал [Annals..., 1988, p. 228]. Подход фон Неймана, напротив, был им хорошо известен (Дж. Маккарти, М. Минский и, возможно, А. Ньюэлл слушали его лекции в Принстонском университете). Фон-неймановская точка зрения, однако, была неприемлема, ибо закрывала путь к искусственному интеллекту, а эта идея, пробужденная кибернетикой, уже завладела умами.

Единственный позитивный путь, который могла предложить тогда кибернетика — моделирование нейронных сетей человеческого мозга, ведущее начало от статьи американского нейрофизиолога У. Маккалока и логика У. Питтса "Логическое исчисление идей, относящихся к нервной активности" 1943 г. [Маккалок и др., 1956]. В этой классической работе была введена формальная модель нервной клетки. Маккалок и Питтс представили нейрон в виде логического элемента с несколькими возбуждающими и тормозящими входами (аналогами синапсов) и одним выходом. Возбуждение выхода (по принципу "все или ничего") происходило при превышении входными возбуждениями некоторого порога и отсутствии торможений. Нейроны соединялись в сети, описываемые на языке логики Карнапа с дополнительными обозначениями Рассела и Уайтхеда (именно у Рассела и Карнапа учился логике У. Питтс). В статье утверждалось, что "для всякого логического выражения, удовлетворяющего некоторым условиям, можно найти сеть, имеющую описываемое этим выражением поведение" [там же, с. 362].



Модель Маккалока-Питтса представляла собой фиксированную структуру, реализующую определенную, наперед заданную форму поведения. В дальнейшем эта модель была усовершенствована, приобрела случайную структуру и возможность "обучения". Сам Маккалок понимал ее ограниченность и даже сказал однажды: "Не хватайте меня за палец, а смотрите, куда я указываю" [Цит. по: McCorduck, 1979, p. 76]. Основная идея, которую он имел в виду и которая была сохранена при всех дальнейших модификациях — установление связи между нейрофизиологическим и феноменологическим уровнями описания путем синтеза структуры сети в соответствии с заданной феноменологией. Коротко говоря, в схеме Маккалока-Питтса не нейронная сеть рождает мысль, а мысль порождает нейронную сеть. При этом сеть уже не моделирует конкретные структуры человеческого мозга, а является лишь инструментом реализации определенного логического механизма переработки сигналов. Фактически, это уже был отказ от моделирования на физическом уровне, шаг к функциональному подходу искусственного интеллекта.

Подходом Маккалока-Питтса вначале увлеклись многие, в том числе и М. Минский, заканчивавший в начале 1950-х годов Гарвардский университет. В качестве дипломной работы он взялся за постройку обучающейся машины из 400 вакуумных трубок, выполнявших функции нейроноподобных элементов Маккалока-Питтса. Машина должна была моделировать поведение в лабиринте четырех крыс, обучающихся избегать встреч друг с другом. Диплом Минский защитил, но машина так и не заработала. Обучение в случайных нейроноподобных сетях шло медленно и неэффективно. С инженерной точки зрения кибернетическая программа здесь явно буксовала, да и теоретическая модель нейрона вызывала критику у нейрофизиологов. "Мне было скучно заниматься тем, что не работало, " – вспоминает теперь Минский [Цит. по: McCorduck, 1979, p. 86]. Как примерный ученый, отказывающийся от бесплодного направления в науке, он перешел к исследованиям по искусственному интеллекту.

## 2. 1. 2. Зарождение исследовательских программ

"Мы предлагаем провести летом 1956 г. двухмесячные исследования по искусственному интеллекту с участием 10 человек в Дартмутском колледже, Гановер, штат Нью-Гэмпшир. Исследования велись бы на основе предположения, что любой аспект обучения или другого свойства интеллекта может в принципе быть описан настолько строго, что возможно создание имитирующей его машины" [Цит. по: McCorduck, 1979, p. 93]. Эти слова содержались в заявке, направленной в фонд Рокфеллера четырьмя учеными: двумя известными – К. Шенноном и Н. Рочестером, придавшими заявке вес своими именами – и двумя совсем молодыми – Дж. Маккарти и М. Минским. Фонд выделил \$7,500, и на Дартмутский летний исследовательский проект, как официально именовалось это мероприятие, съехалось, помимо четырех организаторов, еще 6 человек: Т. Мор, А. Ньюэлл, Г. Саймон, О. Селфридж, Р. Соломонов и А. Сэмюэль.

"Дартмутскую десятку" составили специалисты по математике, статистике, психологии, инженерии, биологии, лингвистике, теории менеджмента. Объединяла их убежденность в том, что путь к искусственному интеллекту — в создании компьютерных программ, имитирующих интеллект функционально, а не в разработке кибернетических устройств, воспроизводящих структуру человеческого мозга. Последний путь был уже испробован и не привел к успеху. Как заметил Минский, ему "потребовалось много

времени, чтобы перейти от попыток понять, как работает мозг, к пониманию того, что он делает" [Цит. по: McCorduck, 1979, p. 84].

Сходство взглядов участников проекта на этом практически заканчивалось. Далее возникали разногласия, которые не позволили им ужиться в рамках одной исследовательской программы. Выявились четыре принципиально различных подхода к проблеме искусственного интеллекта, составившие содержание "ядер" четырех исследовательских программ в этой области. Мы будем называть эти программы, следуя используемой их адептами терминологии: логистическая, эвристическая, коннекционистская и эволюционная.

"Жестким ядром" логистической исследовательской программы стало представление о мышлении как логическом выводе. Произведенное на рубеже XX века усилиями Э. Гуссерля и Г. Фреге отделение логики от психологических наслоений сделало ее строгой наукой. Предметом логики стали законы, которым подчиняется, истина, а не закономерности часто заблуждающегося человеческого мышления. Дж. Маккарти, основатель логистической программы искусственного интеллекта, попытался создать новую логику как теорию мышления (не обязательно человеческого). Используя, как и Фреге, исчисление предикатов первого порядка, он искал формальные выражения для фактов обыденного мира и правил здравого смысла, позволяющих производить простейшие действия в этом мире.

Маккарти приступил к разработке компьютерной системы Advice Taker ("Советующийся"), избрав для начала задачу описания поведения обезьяны, которая, чтобы достать с высоты бананы, должна встать на ящик [McCarthy, 1968b]. Система должна была анализировать поступающую информацию (включая факты и "советы"), сопоставлять с уже известными фактами и делать вывод, как поступить. Основную проблему Маккарти видел не в обилии релевантных фактов (в столь простых ситуациях их действительно было немного), а в разработке богатой логики, учитывающей модальные, каузальные и иные отношения реального мира. "Единственная причина, по которой мы еще не достигли успеха в моделировании всех аспектов реального мира, – заявил он, – это отсутствие достаточно мощного логического исчисления. Я в настоящее время занимаюсь этой проблемой" [Цит. по: Вейценбаум, 1982, с. 260]. Кредо "логицистов" можно изложить в формулировке В. Рапарота: "Логика – в сердце рассуждения, а рассуждение – в сердце интеллекта" [Raparort, 1987, p. 537].

Установка Маккарти была скорее инженерной – построить систему, эффективно решающую задачи, а не смоделировать человеческое мышление. В одном из интервью он заявил: "Я работаю на уровне психологии компьютера – с программами, а не с людьми" [Дискуссии..., 1970, с. 37]. Маккарти стремился вовсе устранить субъективный момент из системы: "Мне кажется, что механизмы интеллекта имеют объективный характер. Когда человек, машина или марсианин играют в шахматы, чтобы добиться успеха, любой из них должен использовать много различных естественных механизмов. Причем эти механизмы определяются существом задачи, которую надо решать, и не зависят от того, кто решает эту задачу" [там же, с. 39].

В то же время Advice Taker был явно построен по образцу именно человеческого рассуждения. Обращая внимание на императивный характер инструкций в обычных компьютерных программах в отличие от декларативных описаний, привычных для человека, Маккарти строит "Советующегося" как раз на базе декларативных представлений. Одно из главных требований автора к системе – чтобы работать с ней

было не сложнее, чем инструктировать человека [McCarthy, 1968, p. 405]. Наконец, работа с Advice Taker натолкнула Маккарти на некоторые теоретические выводы, касающиеся человеческой психологии. В частности, он предположил, что разделение сознания и бессознательного у человека возникает на границе между выражениями типа "стимул – реакция", которые не обсуждаются, а выполняются, и теми выражениями, которые служат лишь посылками в последующем дедуктивном выводе [Ibid., p. 409]. Несмотря на исходную инженерную установку Маккарти, а может быть, и вопреки ей, разрабатываемая им система искусственного интеллекта объективно функционировала в качестве модели человеческого мышления, ее работа сравнивалась с человеческим рассуждением, а результаты интерпретировались в психологических терминах.

В основном же исследования Маккарти шли в русле разработки логического исчисления, а не решения конкретных задач. В 1958 г. он занялся созданием специального языка программирования для искусственного интеллекта – LISP, позднее участвовал в проектировании системы разделения времени для работы в диалоговом режиме. "Советующийся" тем временем не поднимался выше банальных бананов, доставаемых обезьяной.

Основателями другой, более удачливой – *эвристической* исследовательской программы стали математик Алан Ньюэлл и специалист по теории менеджмента, будущий лауреат Нобелевской премии по экономике (1978 г.) Герберт Саймон, создавшие лабораторию искусственного интеллекта в Технологическом институте Карнеги (позднее – Университет Карнеги-Меллона). К представлениям о человеческом мышлении, бытовавшим в философии и нейрофизиологии, они отнеслись с большим скептицизмом. Ньюэлл заявил в интервью, что если наука сталкивается с философией, философия отстывает. "Не обязательно иметь философское обоснование своей работы. Это – последнее из того, что нужно" [Цит. по: McCorduck, 1979, p. 120-121]. Саймон, в свою очередь, посетовал: "Если мы будем ждать, пока физиологи дадут нам теорию мышления, мы никогда этого не дождемся" [Цит. по: McCorduck, 1979, p. 250].

Между тем желание Ньюэлла и Саймона "демистифицировать мышление" росло. Подходящую возможность предоставил заказ военного ведомства США на разработку системы, моделирующей поведение оператора ПВО за экраном радара. Работая над этой системой, Ньюэлл и Саймон использовали ЭВМ не для цифровых вычислений, а для обработки символьной информации. Тогда у них и возникла мысль, что люди и компьютеры – лишь две разновидности систем символьной обработки, и, следовательно, компьютер – наилучший инструмент моделирования человеческого мышления. В способе же такого моделирования заключалась вся новизна предложенной ими исследовательской программы.

Во время учебы в Стенфордском университете Ньюэлл слушал лекции математика Дж. Пойа, развивавшего идею эвристических принципов мышления [Пойа, 1959, 1975]. Эвристика – чисто человеческий, "нелогичный" путь решения задачи, не обязательно приводящий к успеху, но часто оказывающийся эффективнее строгих алгоритмических методов, тем более в задачах, где такие методы вообще неприменимы. Э. Фейгенбаум и Дж. Фельдман предложили следующее определение: "Эвристика (эвристическое правило, эвристический метод) – это основанное на опыте правило, стратегия, ловкий прием, упрощение или иное средство, существенно ограничивающее поиск решения сложных задач. Эвристика отнюдь не гарантирует оптимальность решений: все, что можно сказать о полезной эвристике, это то, что она предлагает решения, которые большей частью

оказываются достаточно хорошими" [Вычислительные..., 1967, с. 29].

Эвристический принцип Ньюэлл и Саймон дополнили заимствованной из бихевиоризма "лабиринтной моделью", интерпретирующей мышление как поиск пути к цели (целевому состоянию системы) в пространстве возможных состояний. Модель мышления как эвристического поиска составила "жесткое ядро" их исследовательской программы. Для эффективного поиска пути к цели, сокращения перебора возможностей использовались следующие эвристики: метод уменьшения различий, разбиение на подзадачи, планирование, применение оценочных функций и другие.

"Позитивной эвристикой" данной исследовательской программы (И. Лакатос, очевидно, тоже использовал термин Дж. Пойа) было выявление и формализация эвристических правил, используемых людьми. С этой целью Ньюэлл и Саймон применяли метод "протоколирования": испытуемые рассказывали о каждом шаге своего рассуждения при решении предложенных им задач. Например, при доказательстве теорем в исчислении высказываний были выявлены следующие приемы, присущие человеку: разбиение на подцели, стремление уменьшить длину выражений и т. д. На этих эвристиках была основана первая действующая система искусственного интеллекта – Logic Theorist ("Логик-теоретик") [Ньюэлл и др., 1967], распечатки с результатами работы которой Ньюэлл и Саймон привезли летом 1956 г. в Дартмутский колледж. "Логик-теоретик" успешно доказал 28 теорем из труда "Principia Mathematica" Рассела и Уайтхеда, причем доказательство теоремы 2.85 оказалось даже короче и изящнее, чем в книге. Саймон написал об этом Расселу, и тот одобрительно ответил.

Вдохновленные успехом, Ньюэлл и Саймон развили на базе "Логика-теоретика" новую систему – GPS (General Problem Solver – "Универсальный решатель задач"). Она не была замкнута на одной предметной области, а предназначалась для любой задачи, сформулированной в специальном виде (как поиск перехода от начального выражения к целевому конечному выражению при заданных правилах уменьшения различий между выражениями). "Универсальный решатель задач" был опробован на решении логических головоломок [Ньюэлл и др., 1967в]. Ньюэлл и Саймон, вместе с программистом Дж. Шоу, продолжали снабжать систему новыми эвристиками и распространять на более широкий класс задач.

О научной или инженерной ориентации своих исследований достаточно определенно высказались сами авторы системы GPS: "Часто приходится слышать, что следует проводить четкую границу между попыткой выполнить при помощи машины задачи, решаемые человеком, и попыткой моделировать процессы, которые человек действительно использует при решении этих же задач... GPS совмещает оба эти подхода с пользой как для того, так и для другого. Первоначально GPS описывалась как попытка создать программу для решения задач... Но вместе с тем варианты GPS моделируют поведение человека" [Ньюэлл и др., 1967, с. 283].

Третья исследовательская программа получила свое начало в работах ф. Розенблатта, О. Селфриджа, Б. Фэрли и У. Кларка [Розенблатт, 1964, 1965, 1966; Фэрли, 1964; Rosenblatt, 1962; Selfridge, 1959]. В отличие от "лабиринтной модели" эвристического поиска, центральным свойством которой была иерархичность мыслительного процесса, исходная модель данной исследовательской программы опиралась на нейрофизиологические представления о сети нейронов головного мозга как гетерархической структуре.

Первые системы искусственного интеллекта, построенные по типу нейронных сетей,

базировались на гипотезе о том, что нервные клетки мозга вначале практически неотличимы друг от друга и лишь благодаря опыту и ощущениям организуются в сложную целостность, демонстрирующую целенаправленное поведение. Такие системы, в которых в результате обучения происходил переход от случайного, "несвязанного" состояния к "связанному", организованному, стали называть "самоорганизующимися". Все они были схожи по основным принципам работы; правда, "персептрон" Розенблатта был реализован "в металле", тогда как система Фэрли-Кларка и "Пандемониум" Селфриджа моделировались на ЭВМ. Эти системы решали одну и ту же техническую задачу распознавания зрительных образов – печатных букв и простых геометрических фигур. Системы "обучались", настраивая сеть на многочисленных примерах. При этом, естественно, использовались чисто инженерные критерии эффективности, точности распознавания. В то же время Розенблатт, например, считал свою задачу чисто научной: "Персептрон является прежде всего и главным образом моделью мозга, а не устройством для распознавания образов. Его использование в качестве модели мозга позволяет определить физические условия образования различных физиологических свойств" [Розенблатт, 1965, с. 18]. Заметим, что в данном случае научная и инженерная цели вновь предполагают друг друга: для построения модели зрения оказывается необходимым сконструировать самосовершенствующуюся зрительную систему.

Данная исследовательская программа, добившись первых успехов в распознавании отдельных букв, вскоре столкнулась с трудностями. Более сложные случаи требовали громадного числа обучающих примеров, а результат достигался далеко не всегда. «Централизованное управление пассивными элементами в системах Розенблатта и Фэрли-Кларка фактически редуцировало гетерархический принцип к статистической иерархии. Более последовательно проводил гетерархические идеи Селфридж, но его подход требовал машинной памяти и быстрого действия, недоступных ЭВМ того времени. Достоинства гетерархии, связанные с параллелизмом и децентрализацией управления, проявятся гораздо позднее, когда эта исследовательская программа возродится и получит наименование "коннекционизм".

"Жестким ядром" четвертой, *эволюционной* программы явилось представление о "разумной" машине как о "tabula rasa" ("чистой доске"), которая, эволюционируя, достигнет известного уровня интеллекта [Friedberg, 1958; Friedberg et al., 1959]. Не исключено, что основатель эволюционного подхода Р. Фридберг был знаком со статьей Тьюринга 1950 г., в которой тот предлагал построить "машину-ребенка": "Почему бы нам, вместо того, чтобы пытаться создать программу, имитирующую ум взрослого, не попытаться создать программу, которая бы имитировала ребенка? Ведь если ум ребенка получает соответствующее воспитание, он становится умом взрослого человека. Как можно предположить, мозг ребенка в некотором отношении подобен блокноту, который мы покупаем в киоске: совсем небольшой механизм и очень много чистой бумаги. Наш расчет состоит в том, что устройство, ему подобное, может быть легко спрограммировано" [Тьюринг, 1960, с. 52].

Фридберг разработал "Обучающуюся машину" – компьютерную программу, предназначенную для автоматического составления простых вычислительных программ. При этом она должна была научиться составлять их без предварительного объяснения, как это сделать. Единственный путь к этому, по мнению Фридберга – слепой случайный поиск. Заметим, что эвристический поиск, напротив, как раз предусматривает подсказку системе, как ей действовать.

"Обучающаяся машина" начинала с произвольной последовательности машинных команд, применявшихся к исходным данным. Далее система двигалась по градиенту, выбирая направление в зависимости от близости полученного результата к требуемому. На каждом шаге в последовательности изменялась одна команда (аналог мутации в генетике). Если результат был благоприятным, это изменение закреплялось и происходила "мутация" другой команды, в противном случае – возврат на шаг назад. В отличие от предложенной Тьюрингом процедуры, напоминающей онтогенез, Фридберг использовал скорее "филогенетический" механизм, включающий мутации и отбор по результатам деятельности. Неудачная последовательность команд "погибала", а "выжившая" давала "потомство". Фридберга, однако, постигла неудача: эволюция шла медленно и неэффективно, к тому же выяснилось, что наследование свойств "родителей" вообще не имеет смысла, т. к. чисто случайный поиск из того же состояния дает требуемый результат в 1000 раз быстрее!

Итак, на первом этапе развития области искусственного интеллекта зародились четыре исследовательские программы: логистическая, эвристическая, коннекционистская и эволюционная. Необычная постановка проблемы, вызвавшая поток вопросов со стороны ("А может ли машина мыслить?"), потребовала от исследователей быстрой отдачи, необходимости реально продемонстрировать, что машина действительно может вести себя разумно. Поэтому на первом этапе преобладала скорее инженерная постановка вопроса, важен был прежде всего конкретный результат. Но уже тогда ведущие системы (типа "Универсального решателя задач") начинали рассматриваться и как теории человеческого мышления.

## **2. 2. 1960-е годы: смена задач**

### **2. 2. 1. Кризис "традиционного" искусственного интеллекта**

В начале 1960-х годов бурный рост сообщений о впечатляющих успехах искусственного интеллекта несколько притормозился. Ученик Г. Саймона Э. Фейгенбаум в обзоре 1968 г. "Искусственный интеллект: темы исследований во втором десятилетии развития" так характеризует этот период: "Ряд наблюдателей отмечает падение продуктивности исследований в 1960—1963 годах, и оценка эта, по всей вероятности, вполне обоснованна. Я считаю, что это вызвано следующими факторами:

- сдвиг интересов основных исследовательских центров в сторону технических аспектов создания вычислительных устройств (например, работы по созданию вычислительной машины с множественным доступом в МТИ, проект вычислительной системы с разделением времени, осуществлявшийся в Стенфорде под руководством Маккарти);
- необходимость переоценки смысла и значения результатов, полученных в конце пятидесятых годов;
- переключение внимания на такие задачи, разрешение которых требует длительного времени (например, анализ естественных языков, "роботы", проблема представления);
- организация в вузах факультетов вычислительной техники, составление программ, осуществление учебных курсов и т. п. поглотили значительную долю сил людей, творчески работавших в этой области" [Фейгенбаум, 1973, с. 173].

Соглашаясь с Фейгенбаумом (все перечисленные обстоятельства действительно

имели место), мы все же склонны полагать, что для спада исследований в начале 60-х имелись и причины когнитивного плана. Эвристическая исследовательская программа – лидер первых лет – оказалась неспособной на данном этапе расширить круг решаемых интеллектуальных проблем. "Универсальный решатель задач" демонстрировал свою универсальность лишь в "ограниченной области математических головоломок с относительно небольшим множеством состояний и хорошо очерченных формальных правил. Как и большинство ее современников, система GPS функционировала в таком формализованном микромире, где возникающие проблемы (например, задача 'Ханойская башня', задача о миссионерах и людоедах), с точки зрения людей, проблемами не являются" [Форсайт, 1987, с. 11]. Попытка Ньюэлла и Саймона превратить GPS в обучающуюся систему [Ньюэлл и др., 1964] успеха не имела и была предана забвению.

Хотя методы поиска в пространстве состояний продолжали совершенствоваться, становилось ясно, что круг интеллектуальных проблем, разрешимых таким методом, ограничен. Н. Нильсон в начале 1970-х годов уже констатировал: "В последнее время исследования в области искусственного интеллекта в США в какой-то степени отошли от эвристического поиска. Первая причина этого состоит, по-видимому, в том, что методика эвристического поиска уже доведена до такого уровня развития, при котором дальнейшее изучение приемов поиска едва ли может коренным образом повысить их эффективность. Другая, и более важная причина состоит в том, что, как показывает опыт, обобщенные процессы поиска, взятые сами по себе, как правило, недостаточны для решения по-настоящему сложных задач" [Нильсон, 1973, с. 9-10]

В 1967 г. работы с GPS были прекращены. Авторы понимали, что причины кризиса глубоки. Ньюэлл писал: "Сама программа занимает внушительную часть запоминающего устройства машины, и построение новых структур данных в процессе решения задачи быстро исчерпывает остатки памяти. Поэтому GPS предназначена для решения только простых задач, представление которых не слишком запутанно. Несмотря на то, что больший объем запоминающего устройства делал бы более приемлемым расточительное использование машинной памяти, характерное для GPS, тем не менее концептуальные трудности остались бы непреодоленными" [Ernst et al., 1969, p. 45].

Немногим лучше шли дела у приверженцев логистической исследовательской программы. Вначале здесь тоже бурлил энтузиазм. "Искусственный интеллект – это прикладная логика", – любил говорить Н. Нильсон, продолжая дело, начатое Дж. Маккарти, в Стенфордском исследовательском институте [Цит. по: Johnson, 1986, p. 189]. Но использование логики порождало многие трудности. Если количество аксиом увеличивается в два раза, число возможных сочетаний возводится в квадрат, если в три раза — в куб и т. д. Попытка описать реальный мир в необходимых подробностях приводит к комбинаторному взрыву.

Луч надежды мелькнул в 1965 г., «когда логик А. Робинсон предложил метод резолюций – высокоэффективный способ доказательства теорем в исчислении предикатов первого порядка [Робинсон, 1970]. Ученик Маккарти К. Грин попытался применить его при разработке системы автоматического программирования (некоторый аналог задачи, поставленной Р. Фридбергом). Выяснилось, что в этом случае возникает такое же большое пространство поиска, как и в эвристических методах, и не наблюдается никакого преимущества в сравнении, скажем, с GPS. Сам Дж. Маккарти тоже не добился заметного прогресса: в итоговом сборнике 60-х годов "Семантическая обработка информации" перепечатаны лишь его старые работы 1959 и 1963 гг. [McCarthy, 1968a, 1968b]. Даже

лекции Маккарти в Стенфордском университете стали носить столь неясный характер, что студенты прозвали их "таинственный час дядюшки Джона" [McCorduck, 1979, p. 220].

Эволюционная исследовательская программа получила новый импульс развития, который, однако, так и не сдвинул ее с места. Л. Фогель с соавторами подвергли критике "традиционные" подходы к проблеме создания искусственного интеллекта, которые основывались на моделировании человеческого мышления, на подражании "наиболее разумному существу". Они сочли эту точку зрения слишком эгоцентричной: "Хотя человек, безусловно, разумное создание, нет оснований полагать, что он наиболее интеллектуальное существо, которое вообще возможно. Эксперимент природы продолжается, и ввиду ее прежних успехов в создании существ последовательно все более высокой интеллектуальности кажется вполне разумным предположить, что некоторые будущие существа будут обладать значительно большим интеллектом, чем современный человек. В этом свете было бы, пожалуй, полезным заменить процесс моделирования человека моделированием процесса его эволюции" [Фогель и др., 1969, с. 28]. Недостатки системы, разработанной под руководством Фогеля, оказались теми же, что и у "Обучающейся машины" Р. Фридберга: подъем по градиенту путем случайных мутаций был слишком медленным, застревающим на локальных максимумах. Инженерная задача, поставленная Фогелем (не воспроизвести, а превзойти человеческое мышление), осталась невыполненной. Попытка Х. Бремермана применить для решения системы линейных уравнений процедуры эволюции (включая как "брачную игру", так и бесполой вариант), по признанию автора, также не дала никаких примечательных результатов [Bremermann, 1962, p. 101].

Коннекционистская программа породила прикладную область создания устройств распознавания образов, далее развивающуюся уже по линии технического совершенствования. В теоретической же сфере подход нейронных сетей подвергся уничтожающей критике в нашумевшей книге М. Минского и С. Пейперта 1969 г. "Перцептроны" [Минский и др., 1971]. В ней авторы показали, например, что перцептрон, умея распознавать буквы А и В, комбинацию ВА воспринимает как новый, совершенно незнакомый объект. Отсутствие символьных представлений не позволяло перцептрон использовать в явной форме накопленный опыт. Ранняя смерть создателя перцептрона Ф. Розенблатта и отсутствие достойного ответа коннекционистов на критику привели к тому, что большинство научного сообщества отвернулось от этого подхода. В действительности свою неадекватность продемонстрировал лишь первый "защитный пояс" исследовательской программы, но не его "жесткое ядро". Позднее Пейперт признал, что в книге "Перцептроны" на самом деле "доказывались серьезные ограничения определенного класса сетей (сейчас называемых однослойными перцептронами), но ошибочно утверждалось, что это является ядром коннекционизма" [Papert, 1988, p. 4].

Общим недостатком всех действовавших исследовательских программ был выбор интеллектуальных проблем, далеких от реальных задач, встающих перед человеком в обыденной и профессиональной сферах. Системы искусственного интеллекта успешно решали лишь искусственные задачи типа логических головоломок. Один из основоположников когнитивной психологии У. Найссер отмечал в начале 70-х: "Изучение процессов переработки информации становится все более распространенным и престижным, однако оно еще не связано с такой теорией человеческой природы, которая могла бы найти себе применение за пределами лаборатории" [Найссер, 1981, с. 28]. Из приведенного А. Ньюэллом списка 59 феноменов, моделируемых на компьютерах, лишь



два обладают "экологической валидностью" – игра в шахматы и наблюдение за Луной [Newell, 1973б].

Наконец, последний фактор, определивший регресс четырех "традиционных" исследовательских программ – возникновение новой, *когнитивной* программы, сразу вырвавшейся вперед.

## 2. 2. 2. Искусственный интеллект в естественных условиях

Когнитивная исследовательская программа обязана своим появлением на свет ряду ученых, но прежде всего – профессору МТИ М. Минскому. Как мы помним, не в его характере "заниматься тем, что не работает". Минский известен в научном сообществе широтой своих взглядов и восприимчивостью к новым идеям (недаром первый обзор исследований по искусственному интеллекту принадлежит именно ему [Минский, 1967б]). И в этой кризисной ситуации Минский сумел найти выход.

Прежде он исповедовал логико-алгебраический подход и употреблял формулировки типа "мышление представляет собой вид эффективного вычисления" [Минский, 1971, с. 351]. Э. Хант писал, что Маккарти, Минский и их коллеги "рассматривали искусственный интеллект скорее как расширение математики и символической логики, чем как дисциплину, параллельную психологии" [Хант, 1978, с. 18]. Но с уходом Маккарти в 1963 г. из МТИ в Стенфордский университет в их взглядах начинают все более ощущаться различия. Позже Минский облек свое мнение о работах Маккарти в форму каламбура: "Математика должна быть на углу стола, а не во главе угла" [Цит. по: McCorduck, 1979, p. 335] (английская фраза "Mathematics should be on tap but not on top" буквально переводится "Математика должна быть под рукой, а не на вершине").

Минский, по-видимому, раньше других почувствовал кризис и начал искать новый подход. В статье 1961 г. он писал: "Исследования по искусственному интеллекту развиваются медленно. Трудности носят концептуальный характер: методы, хорошо работающие на легких задачах, не переносятся беспрепятственно на трудные. Необходимо использовать нетривиальные формальные и описательные языковые системы... следует включить более сложные лингвистические процессы..." [Minsky, 1968a, p. 419]. Для распространения возможностей интеллектуальных систем на реальные задачи, с которыми сталкивается человек в естественных условиях, а не в стерильном интеллектуальном пространстве логических рассуждений, потребовались иные образцы интеллектуальной деятельности. В поисках новой исходной модели мышления Минский обращается к лингвистике и психологии зрения. Суть "жесткого ядра" новой исследовательской программы составляет метафора "мышление – это познание". Ее "позитивная эвристика" состоит в моделировании человеческих способов представления и семантической обработки знаний.

В 1960-е годы под руководством Минского в МТИ было защищено несколько диссертаций, в которых намечался новый подход к искусственному интеллекту. Эти работы вошли в сборник "Семантическая обработка информации" [Semantic..., 1968]. Там впервые были предложены способы представления информации, моделирующие человеческие представления. Автор программы организации памяти в виде семантической сети Р. Квиллиан писал, что в его программе воплощена "разумная точка зрения на то, как организована семантическая информация в мозгу человека" [Quillian, 1968, p. 217]. Значения отдельных слов и словосочетаний хранились в памяти в форме сети, состоящей

из вершин (понятий) и соединяющих дуг (отношений). Ученик Минского Д. Бобров. написал программу STUDENT, решающую алгебраические задачи на естественном языке, предварительно переводя их в форму уравнений. По мнению автора, разработанная им "семантическая теория словесного общения может послужить основой для значительно более общей системы переработки языковых соотношений" [Bobrow, 1968, p. 194].

В задачах понимания языка и визуального восприятия оказалось много общего, т. к. и те, и другие касаются знания и его представления, обучения понятиям, установления правил распознавания. Комплекс лингвистико-визуальных исследований в МТИ получил название "Проект MAC", который расшифровывают как "Man and Computer" ("Человек и компьютер"), а иногда как "Machine-Aided Cognition" ("Познание с помощью машины").

Наибольшие успехи когнитивной исследовательской программы связаны с разработкой Плэннера (PLANNER), нового языка программирования для систем искусственного интеллекта. Его автор К. Хьюитт, также профессор МТИ, подверг критике эвристический и логистический подходы. Эвристический поиск в пространстве состояний требует, чтобы были строго определены: начальное состояние (например, исходная позиция в шахматах), процедуры (разрешенные ходы) и конечные (целевые) состояния (мат, пат или вечный шах). В реальных ситуациях человеческой деятельности эти характеристики нельзя определить достаточно строгим и исчерпывающим образом. Пример, приведенный Хьюиттом – освоение Северной Америки европейцами, когда начальное состояние было толком неизвестно, процедуры импровизировались на ходу, а цели менялись во времени. "Решение задач в открытых системах, – пишет Хьюитт, подразумеваемая под "открытой" современную компьютерную систему, взаимодействующую с внешним миром, – больше похоже на процесс освоения Америки, чем на игру в шахматы" [Хьюитт, 1987, с. 91].

Уязвимым оказывается и логистический подход. "Точка зрения логиков на противоречивые теории ясна: эти теории бессмысленны, поскольку не соотносятся ни с одним из 'возможных миров'. Логическая трактовка мира слишком узка для сложных реальных систем, так как противоречивые знания и описания не бессмысленны" [там же, с. 95]. Напротив, "любая система аксиом, пытающаяся выразить знания человека о мире, неизбежно оказывается противоречивой" [там же, с. 94].

В разработанном в 1969 г. языке Плэннер Хьюитт стремился преодолеть эти трудности. Он обеспечил решение задач поиска по цели, не требуя при этом отдельного описания пространства состояний. При наличии конфликтной информации и противоречивых фактов Хьюитт предложил использовать вместо логического вывода "процессуальное рассуждение", в ходе которого анализируются свидетельства и аргументы в пользу различных точек зрения и вырабатывается некоторое соглашение.

Т. Виноград, ученик Минского в МТИ, на рубеже 70-х гг. написал на Плэннере программу SHRDLU, которая умела вести беседу на ограниченном естественном языке по телетайпу и одновременно работать с компьютерным представлением "микромра" (простой сцены, включающей несколько кубиков, пирамид и других фигур); Программа анализировала ситуации в "микромра", вносила изменения, отвечала на вопросы [Виноград, 1976]. Система успешно выполняла простые команды типа "поставь пирамиду на куб", разрешая синтаксическую неоднозначность входных фраз за счет использования семантической информации, полученной в результате "визуального" анализа "микромра". Знания в SHRDLU не составляли логически стройной системы определений и аксиом. Скорее это была совокупность понятий, отражающая сложность и запутанность

человеческих знаний. Определения понятий, по выражению автора, "циркулярны", т.е. циклически связаны между собой: значение каждого из них зависит от других понятий.

При создании SHRDLU Виноград указал тройкую цель:

- 1 – осуществить на практике удобную систему, понимающую язык;
- 2 – лучше понять, что такое язык и как он устроен;
- 3 – понять, что такое интеллект и как он может быть представлен в машине [там же, с. 11].

Научный и инженерный аспекты проблемы вновь оказались неразрывно связанными. Разработка системы, решающей реальные задачи, явилась средством познания механизмов интеллектуальной деятельности. Исследование интеллекта, в свою очередь, помогло построить эффективную компьютерную систему.

Главным теоретическим результатом работ с SHRDLU явился вывод о единстве всей системы человеческих знаний о мире, используемой в любом, даже частном интеллектуальном акте. Как сказал Г. Хендрикс из Стенфордского исследовательского института, "естественная языковая коммуникация – это деятельность всего интеллекта" [Цит. по: Johnson, 1986, p. 118]. Этот важный вывод, полученный когнитивной программой, привел к использованию в исследованиях по пониманию естественного языка и машинному переводу экстралингвистического знания (о целях, намерениях, убеждениях говорящего), названного "прагматикой" языка.

Подытоживая картину развития области искусственного интеллекта в 60-е годы, отметим теоретический застой в большинстве исследовательских программ при постепенном совершенствовании инженерной техники. Прежние теоретические модели оказались бессильными перед реальными, а не искусственными интеллектуальными задачами. Успеха в освоении этой сферы добилась лишь новая, когнитивная программа, в рамках которой мышление уподобляется пониманию языка или зрительному восприятию и требует привлечения всей системы знаний о мире, а не только правил формальной игры.

## **2. 3. 1970-е годы: конкуренция исследовательских программ**

### **2. 3. 1. Проект DARPA**

К началу 70-х годов методы искусственного интеллекта уже работали во многих областях, в том числе применялись и для задачи понимания речи. Инженерные параметры типичной системы, разработанной в Университете Карнеги-Меллона, были таковы: голос фиксированного пользователя; узкая предметная область ("микромир" кубиков); всего 16 слов, которые по искусственным синтаксическим правилам могли образовывать 192 предложения. При точности в 85% система требовала на понимание в 10 раз больше времени, чем уходило на произнесение фразы.

В проблеме понимания речи сконцентрировалось множество важных и интересных когнитивных и инженерных задач, привлекавших как исследователей искусственного интеллекта, так и заказчиков извне. В 1971 г. Управление высших исследований Министерства обороны США (The Defense Department's Advanced Research Projects Agency (DARPA)) объявило конкурс на создание лучшей системы понимания речи. В качестве приза гарантировалось финансирование победившего проекта в течение 10 лет. Хотя целью провозглашалось стимулирование фундаментальных исследований, требования были предъявлены инженерные, и притом самые жесткие. За 5 лет необходимо было

сконструировать систему, не ориентированную на фиксированного пользователя (требовался лишь стандартный американский диалект) и понимающую фразы, использующие 1000 слов из ограниченной предметной области (однако, пошире, чем "микромир" кубиков).

Для того, чтобы удовлетворить этим условиям, в области понимания речи требовался подлинный прорыв. Пять лабораторий искусственного интеллекта приняли вызов. В соревновании приняли участие специалисты из Университета Карнеги-Меллона (группа под руководством Р. Редди), фирмы "Болт, Беранек и Ньюмен" (под руководством В. Вудса), Стенфордского исследовательского института (под руководством Н. Нильсона), Линкольновской лаборатории МТИ и Корпорации развития систем. Ф. Хейес-Рот, сотрудник Редди, так описывал ситуацию: "Не только DARPA, весь мир смотрел на нас. Это было испытание боем. Так было объявлено, так ученые и взялись за это. В данный момент определялось, имеет ли искусственный интеллект право на существование" [Цит. по: Johnson, 1986, p. 130].

Вскоре группа из МТИ прекратила работы, а Стенфордский исследовательский институт и Корпорация развития систем объединили свои усилия в едином проекте. Три оставшиеся конкурсанта представляли три различные исследовательские программы: эвристическую (Университет Карнеги-Меллона), когнитивную ("Болт, Беранек и Ньюмен") и логистическую (Стенфордский исследовательский институт и Корпорация развития систем). Это был первый случай явной конкуренции всех основных исследовательских программ искусственного интеллекта, причем протекающей по очень жестким правилам.

Группа Редди подошла к проблеме с чисто инженерной стороны. Их не интересовал реальный вклад фонетических, лексических, синтаксических, семантических и прагматических знаний в процесс человеческого понимания речи. Соотношение между этими компонентами решено было установить эмпирически – с помощью специальной процедуры, названной "blackboard" (иногда переводится как "школьная доска", но более точный смысл – "доска объявлений"). Созданная в Университете Карнеги-Меллона система HEARSAY-II воспроизводила работу группы консультантов (где каждый имеет собственный набор знаний), обменивающихся гипотезами и результатами их проверок через общее поле — "доску объявлений". Информация на "доске" появлялась обезличенной; не было единого управляющего органа, и каждый консультант преследовал собственную цель. При этом автоматически задействовались наиболее полезные знания, способные проверить попавшие на "доску" гипотезы и, при необходимости, выдвинуть новые. Знания в системе представлялись в форме эвристик вида "ЕСЛИ условие ТО действие". "Условие" в данном случае — определенный набор информации на "доске объявлений", "действие" – модификация или создание новых элементов на той же "доске". Система HEARSAY-II мало использовала семантику и совсем не привлекала прагматических знаний, обрабатывая лишь предложения ограниченного числа типов [Erman et al., 1980].

Группа Вудса поставила перед собой гораздо более общую задачу. Рассматривался широкий класс предложений, учитывался не только синтаксис, но и семантика, прагматика и просодика (интонация). Система HWIM ("Hear What I Mean" ("Слушай, что я хочу сказать")) следовала созданной Вудсом лингвистической модели "расширенной сети переходов" [Вудс, 1978], в которой понимание предложения — это процедура, оперирующая семантическими примитивами ("атомами значения"). При создании

системы некоторые ее части вначале заменяли люди, и лишь постепенно их место занимали компьютерные программы. Для Вудса задача познания процесса человеческого понимания речи явно преобладала над инженерными целями. Но эта задача оказалась слишком трудной для пятилетнего срока конкурса. За последние несколько месяцев надо было добавить в систему около 400 слов до требуемой тысячи. Времени на точную настройку подсистем уже не хватило.

Третий проект не был закончен в срок. Корпорация развития систем разрабатывала нижние уровни анализа, а Стенфордский исследовательский институт – верхние. Состыковать эти части в единое целое оказалось делом слишком трудным и долгим.

В 1976 г. было проведено испытание, выявившее победителя. На контрольных примерах система HWM достигла точности 44%, HEARSAY-II – 91%. Победила же совершенно неожиданно "темная лошадка" – система HARPУ, созданная Р. Редди и Б. Лоуэрром в свободное от основных занятий время. Система не использовала никаких знаний высокого уровня, а лишь строила сеть, хранящую шаблоны всех допустимых в данном случае предложений. Понимание сводилось к эвристическому поиску подходящего шаблона.

Вудс справедливо усмотрел причины своего проигрыша в расхождении научных и инженерных целей разных исследовательских групп: "Проекты фирмы 'Болт, Беранек и Ньюмен' и Университета Карнеги-Меллона – почти полярные подходы к пониманию речи. Система HARPУ буквально удовлетворила требованиям, но сделала это таким путем, который более никуда не ведет" [Цит. по: Johnson, 1986, p. 138-139]. Даже те, кто восхищался инженерной изощренностью HARPУ, признавали, что эта система отвечает букве, а не духу условий DARPA. Ф. Хейес-Рот, работавший над HEARSAY-II, дал версию победившей стороны: "В мире искусственного интеллекта Университет Карнеги-Меллона считается местом, где сидят инженеры. У меня нет почти никаких претензии на вклад в лингвистику; я был заинтересован лишь в развитии инженерии знаний в области искусственного интеллекта. Все мы разные, но, в сущности, заурядные специалисты по речи. Прежде всего мы – специалисты по компьютерам. Для нас представляет интерес та часть искусственного интеллекта, которая касается создания работающих систем, а не глубокие теоретические изыскания. Фирма 'Болт, Беранек и Ньюмен' сделала великолепную работу в области науки о речи, но они поставили проблемы, которые не смогли решить" [Цит. по: Johnson, 1986, p. 139]. Группа из Университета Карнеги-Меллона получила весь обещанный DARPA приз; остальные не получили ничего.

Расхождение научных и инженерных критериев оценки исследований стало сказываться очень резко. Прогресс, достигнутый когнитивной исследовательской программой в построении теории человеческого понимания речи, обернулся поражением в инженерии. Прежнее убеждение многих специалистов по искусственному интеллекту, что именно изучение человеческого мышления позволит конструировать более эффективные интеллектуальные компьютерные системы, начало быстро рассеиваться.

### **2. 3. 2. Раскол научного сообщества**

Вплоть до начала 70-х годов многие специалисты по искусственному интеллекту считали, что научные и инженерные цели исследований вполне согласуются друг с другом. Неудачи в работе систем объяснялись еще недостаточной изученностью процессов человеческого мышления, при этом предполагалось, что более глубокое их

познание и более точное моделирование позволят достичь и инженерного успеха. В начале 70-х, благодаря усилиям Х. Дрейфуса [Дрейфус, 1978] и других философов, стало ясно, что далеко не все аспекты интеллектуальной деятельности поддаются компьютерному моделированию (в частности интуиция, инсайт). С другой стороны, становилось очевидным, что человек отнюдь не во всех ситуациях действует наилучшим образом, так что более точное моделирование может даже снизить эффективность системы. Научная и инженерная цели стали противоречить друг другу. В этой ситуации в научном сообществе произошел раскол на две школы: "интеллектуальные артефакты" и "когнитивное моделирование".

В возникновении школы "интеллектуальных артефактов" ведущую роль сыграл Э. Фейгенбаум, бывший ученик Г. Саймона. Фейгенбаум взялся за проблему, открытую когнитивной программой – представление знаний – но подошел к ней с инженерной стороны. Совместно с известным химиком Дж. Ледербергом он разработал первую экспертную систему DENDRAL, предназначенную для синтеза конфигураций химических молекул по данным масс-спектрографии [Feigenbaum, 1971]. Система была основана на работе эвристик — правил такого синтеза, полученных в ходе консультаций с экспертом, в роли которого выступал Ледерберг. DENDRAL показывал результаты не хуже, чем научный персонал лаборатории, а иногда и превосходил людей. В эвристической исследовательской программе появился новый "защитный пояс": на смену универсальным эвристикам, не ориентированным на какую-то конкретную предметную область, пришли специализированные эвристики – знания из узкой профессиональной сферы. Фейгенбаум даже ввел новый термин "инженерия знаний", чтобы отмежеваться от прежнего искусственного интеллекта, претендующего на статус науки. Единственным критерием успеха в школе "интеллектуальных артефактов" стала эффективность работы экспертной системы.

Представители "традиционного" крыла эвристической исследовательской программы – А. Ньюэлл и Г. Саймон – тоже сконцентрировались на проблеме представления знаний. Если прежде они уделяли основное внимание разработке более мощных процедур – методов эвристического поиска – то ныне, подчеркивая роль знаний, Саймон перефразировал известный афоризм Канта: "Знания без подходящих процедур немы, процедуры без соответствующего знания слепы" [Цит. по: McCorduck, 1979, p. 261]. Ньюэлл, в свою очередь, предложил процедурную форму хранения знаний в виде правил-продукций "ЕСЛИ условие ТО действие" [Newell, 1973a], примененных, как описано выше, в системе понимания речи HEARSAY-II. При этом, оставаясь верными научной ориентации исследований, Ньюэлл и Саймон ссылались в качестве обоснования на человеческий способ решения задач [Newell et al., 1972].

Зг

Правила-продукции были использованы Э. Шортлифом в экспертной системе MYCIN для представления медицинских знаний по диагностике инфекционных заболеваний [Shortliffe, 1976], а затем и в ряде других систем прагматической направленности. Новый "защитный пояс" эвристической исследовательской программы продемонстрировал свою эффективность и при решении старых задач. М. Д. Риченер построил несколько продукционных систем, воспроизводящих ранние работы по искусственному интеллекту: EPAM, GPS, SHRDLU, STUDENT.

Логистическая исследовательская программа также приступила к решению реальных, а не искусственных задач. Язык логического программирования PROLOG, разработанный во Франции в начале 1970-х годов, позволил использовать исчисление

предикатов в качестве способа представления профессиональных знаний и писать на нем экспертные системы. Дж. Маккарти приступил к разработке "немонотонной логики", позволяющей распространить логические средства на область описания реального мира, где строгие соотношения не всегда соблюдаются [McCarthy, 1979, 1980]. Он ввел понятие "исключения из правил", позволяющее включать в базу данных новые факты, не вызывая противоречий (все птицы летают; страус – птица; страус не летает; страус – исключение). Обычный вывод монотонен (добавление нового факта не изменяет следствий из прежних фактов), немонотонный же вывод требует при добавлении нового утверждения (например, "страус — исключение"), пересмотра всех следствий. Немонотонная логика активно разрабатывалась в СССР группой под руководством В. К. Финна [Финн, 1983]. Немонотонность изменяет само понятие логического вывода, делает его более приближенным к неформальному человеческому рассуждению.

Школу "когнитивного моделирования" составили в основном приверженцы когнитивной исследовательской программы. К их достижениям следует отнести понятия "фрейма", "плана" и "сценария", являющиеся моделями хранения и использования человеческого опыта.

"Фрейм" был описан М. Минским в работе 1974 г.: "Отправным моментом для данной теории служит тот факт, что человек, пытаясь познать новую для себя ситуацию или по-новому взглянуть на уже привычные вещи, выбирает из своей памяти некоторую структуру данных (образ), называемую нами фреймом, с таким расчетом, чтобы путем изменения в ней отдельных деталей сделать ее пригодной для понимания более широкого класса явлений и процессов" [Минский, 1979, с. 7]. Концепция фрейма основана на исследованиях по компьютерной лингвистике и машинному зрению, ведущихся в МТИ. Типичным фреймом является, например, сцена, изображающая обстановку комнаты: Комната (Пол (Цвет, Материал), Потолок, Стены (Цвет), Мебель, Окно). "Слоты" типа "Мебель" или "Цвет пола" уточняются при восприятии конкретной комнаты. В самой структуре фрейма закладывается человеческий опыт восприятия стереотипных ситуаций, группировки признаков по привычным "отсекам".

Понятия "план" и "сценарий" были введены группой исследователей из Йельского университета во главе с Р. Шенком. Это, по сути дела, некие операциональные аналоги фрейма. Они описывают обыденные события, например, посещение магазина. "Сценарий – это предопределенная стереотипная последовательность действий, которые описывают хорошо известную ситуацию" Шенк и др., 1975, с. 210]. Объекты в сценарии могут получать конкретное значение без его специального введения, т. е. сам сценарий неявно вводит их. "Планы описывают множество выборов, которые есть у лица, когда оно стремится к достижению цели. Слушая повествование, люди используют планы, чтобы понять предложения, которые кажутся совершенно не связанными друг с другом. Обнаружив некий план, "понимающий" может строить догадки относительно намерений в действии развертывающейся истории и использовать эти догадки для осмысления истории" [там же, с. 216].

Фреймы, планы и сценарии позволяют выявить ту неявную информацию, которая при обычных описаниях опускается именно в силу своей обыденности и общеизвестности. Такой анализ структур человеческого мышления на деле является построением психологической теории компьютерными средствами. Как заявил в интервью Р. Шенк, "граница между искусственным интеллектом и психологией размывается" [Цит. по: Johnson, 1986, p. 179].

В результате произошло расслоение научного сообщества искусственного интеллекта. От основной группы исследователей, по-прежнему гонящихся одновременно за зайцами науки и инженерии, отделились "верхний" и "нижний" пласты. Одни занялись "инженерией знаний", явно устранив постановку задачи моделирования человеческого мышления. Другие выделились в "когнитивную науку". С 1977 г. начал выходить международный журнал "Cognitive Science", редколлегия которого объявила, что журнал не принимает статьи, посвященные лабораторным, а не реальным интеллектуальным задачам и эффективным алгоритмам, не моделирующим человеческое поведение [Cognitive..., 1977].

На наш взгляд, этот раскол все же произошел лишь в рефлексии научного сообщества, а не в самой исследовательской области. Сами по себе модели интеллектуальной деятельности, разработанные и в школе "интеллектуальных артефактов", и адептами "когнитивного моделирования" сохранили свою потенциальную двойственность, возможность включаться как в научный, так и в инженерный контекст. На основе фреймов начали разрабатываться экспертные системы прагматического назначения [Stefik, 1981a, 1981б]. В то же время работы по экспертным системам даже у таких "чистых" инженеров, как Фейгенбаум, были связаны с выявлением, представлением, согласованием и использованием сугубо человеческих экспертных знаний. "Эвристические приемы..., принципы составления... точной теоретической модели... – все это было почерпнуто у наших коллег-химиков при работе в режиме взаимодействия человек-машина" [Фейгенбаум, 197, с. 194]. При этом, даже по инженерным соображениям, приходилось выбирать наиболее привычные и удобные для человека представления знаний. "Невозможно было поступить иначе, – рассказывал в интервью Фейгенбаум, – ибо тогда мы не смогли бы понять, что делается в программе и модифицировать ее" [Цит. по: McCorduck, 1979, р. 283]. Явно указал на познавательный смысл прикладных систем искусственного интеллекта и исследователь человеческого мозга М. Арбиб. Он отметил, что исследование функционирования систем HEARSAY и VISIONS как "множества взаимодействующих источников знаний для анализа акустических сигналов или цветных фотографий может быть полезно для понимания того, как взаимодействуют различные области мозга" [Arbib, 1987, р. 283].

В целом в 70-е годы все три основные исследовательские программы – эвристическая, логистическая и когнитивная – прогрессировали, охватывая новый круг задач, связанных с узкопрофессиональной экспертизой. Теоретические новинки – правила-продукции, немонотонная логика, фреймы, планы и сценарии – позволили им адаптироваться к реальным проблемам, решаемым человеческим интеллектом. Последователи коннекционизма и эволюционного подхода отошли на периферию научного сообщества, не сумев достойно ответить на этот вызов.

## **2. 4. 1980-е годы: наука или коммерция?**

### **2. 4. 1. Инженерный бум**

В начале 80-х годов коммерческий интерес к изделиям искусственного интеллекта резко возрос. Этому способствовал целый ряд факторов. Фирмам, производящим персональные компьютеры, удалось снизить их себестоимость настолько, что они стали доступны для покупки в личное пользование, а также укладывались в сметы расходов



учреждений на канцелярские нужды. Десятки миллионов персональных ЭВМ хлынули в офисы фирм и частные дома. Это вызвало резкий подъем спроса на следующие виды программного обеспечения:

1 – экспертные системы, тиражирующие уникальный опыт лучших специалистов. Медицинскую, агрономическую, юридическую консультацию стало возможным получить в самом отдаленном уголке, подключившись к компьютерной сети или взяв дискету с собой на борт корабля или самолета;

2 – системы общения на естественном языке, облегчающие работу с компьютером пользователю, не являющемуся профессиональным программистом;

3 – системы автоматического программирования, в несколько раз повышающие производительность программистского труда;

4 – интеллектуальные игры для детей и взрослых;

5 – обучающие системы по самым различным предметам;

6 – справочные и информационно-поисковые системы, обеспечивающие доступ к общенациональным банкам данных через компьютерную сеть;

7 – средства коммуникации, обеспечивающие издание компьютерных журналов и проведение заочных конференций.

Основу для разработок в этих областях составили способы представления данных и знаний и методы их обработки, созданные в рамках исследований по искусственному интеллекту.

Провозглашенный в 1979 г. японским Научно-исследовательским центром по обработке информации JJPDEC план развертывания, научно-исследовательских и опытно-конструкторских работ по созданию ЭВМ нового, пятого поколения предусматривал в качестве одной из центральных задач "интеллектуализацию" ЭВМ [Симонс, 1985; ЭВМ..., 1984; Feigenbaum et al., 1983]. В ответ в США и в Западной Европе были также произведены крупные финансовые вложения в сферу информатики. В среднем в 80-е гг. годовой прирост вложенных средств распределялся по отраслям искусственного интеллекта следующим образом [Кузин, 1985, с. 4]:

Отрасль	Годовой прирост, %
Экспертные системы	86
Системы общения на естественном языке	79
Автоматизированное обучение	46
Распознавание визуальных образов	60

Распознавание речи	56
Аппаратные средства	54
Рынок искусственного интеллекта в целом	68

Безусловно, коммерческий спрос предъявлял к системам искусственного интеллекта прежде всего инженерные требования. Потребителей интересовала эффективность работы компьютерных программ, а не их познавательная ценность.

Область искусственного интеллекта, со своей стороны, была готова ответить на эти запросы. Технология создания экспертных систем была уже в основном разработана, по другим направлениям был накоплен достаточный теоретический задел для выпуска коммерческих продуктов. С невероятной скоростью начали плодиться фирмы, производящие интеллектуальное программное обеспечение. И здесь успех стал оборачиваться к искусственному интеллекту своей обратной стороной. В эти фирмы стали уходить ведущие ученые области: У. Вудс, М. Минский, Э. Фейгенбаум, Р. Шенк и многие другие. 25-30 исследователей, ежегодно получающих в США ученую степень по специальности "Искусственный интеллект", стали расхватываться фирмами, не доходя до университетских лабораторий. Р. Гренджер, который в середине 70-х годов занялся искусственным интеллектом, пленившись его романтикой и отказавшись от карьеры бизнесмена, признался, что "сейчас за консультации по искусственному интеллекту получает больше, чем если бы окончил школу бизнеса" [Цит. по: Johnson, 1986, p. 26]. Доля предпринимателей среди участников конференций и симпозиумов стала превышать число специалистов. Если в 1980 г. на 1 сессии Американской ассоциации искусственного интеллекта присутствовало 800 человек, в основном исследователи и студенты, то в 1984 г. участников было уже 3000, две трети из которых не принадлежали к академическому сообществу.

Выступая на сессии 1984 г., Р. Шенк представил и себя, и все научное сообщество в целом таким научно-инженерным кентавром: "Мы так много обещали, что возложили на искусственный интеллект обязанность производить работающие системы. Многие из вас должны посвятить этому себя, и часть меня самого посвящена этому. Другая часть из вас предпочла бы заниматься наукой, и часть меня самого предпочла бы это" [Цит. по: Johnson, 1986, p. 28]. Наверно, такое разделение происходило в душе многих исследователей. Инженерные цели, подкрепленные коммерческим интересом, стали вытеснять научную ориентацию трех основных исследовательских программ – эвристической, логистической и когнитивной. Так и произошло. 1980-е годы стали временем научного прогресса коннекционистской программы.

#### **2. 4. 2. "Булевская" и "небулевская" исследовательские программы**

Наиболее "почтенные", "классические" исследовательские программы искусственного интеллекта – эвристическая, логистическая и когнитивная – сходны в том

отношении, что интеллектуальная деятельность предстает в них в виде последовательного оперирования дискретными символами по определенным правилам, "законам мысли", как выразился Джордж Буль. В этом смысле все они могут быть названы "булевскими". Виднейший представитель эвристической программы Г. Саймон, в частности, так объясняет это убеждение: "Наши рациональные, сознательные процессы выполняются последовательно. Мы можем охватить небольшое количество информации и поэтому разлагаем проблему на части и решаем их одну за другой... Я не вижу никаких оснований считать, что человек при решении задач делает что-то параллельно. Сейчас мои коллеги это активно обсуждают. Но они неправы" [Цит. по: Johnson, 1986, p. 151].

Важной предпосылкой возможности автоматизации мышления служит представление о простоте и немногочисленности "законов мысли". Г. Саймон иллюстрирует этот тезис на примере муравья: "В том, что касается принципов своего поведения, муравей весьма прост. Кажущаяся сложность его поведения во времени в основном отражает сложность внешней среды, в которой он функционирует" [Саймон, 1972, с. 35]. Далее Саймон дословно повторяет данное утверждение, заменяя слово "муравей" на "человек". По сути дела, это есть исходная метафора для любого булевского подхода к искусственному интеллекту. Назовем ее "метафорой муравья".

Коннекционизм противостоит всем этим трем программам благодаря своему небулевскому характеру. Д. Хофстадтер в работе "Пробуждение от булевского сна" так формулирует основную идею: "Мозг сам не манипулирует символами, он – посредник, с помощью которого символы двигаются и приводят в движение друг друга. Нет никакого центрального манипулятора, центральной программы. Есть просто огромное собрание "отрядов" – нейронных сетей, которые, как отряды муравьев, запускают другие нейронные сети... Мы не только не манипулируем символами, фактически наоборот, наши символы манипулируют нами!" [Hofstadter, 1985, p. 648]. Тот же пример муравья для Хофстадтера играет совсем иную роль, чем для Саймона. В книге "Гедель, Эшер, Бах: вечное золотое плетение", ставшей общепризнанным бестселлером искусственного интеллекта [Hofstadter, 1981], Хофстадтер уподобляет мышление муравейнику, в котором простое, "неинтеллектуальное" поведение каждого отдельного муравья рождает разум на уровне их сообщества. В небулевской модели мыслящее целое возникает в результате локальных взаимодействий неинтеллектуальных частей. Мы будем называть это "метафорой муравейника".

В 1982 г. Хофстадтер приступил к разработке на основе этой модели системы LUMBO, предназначенной для решения анаграмм. Система разлагала исходное слово на буквы и комбинировала их в локальные сочетания. Наиболее естественные, присущие языку сочетания образовывали устойчивые элементы, слагающие затем искомое слово. Хофстадтер назвал такой метод "статистически возникающим поведением", когда путь решения не указывается сверху, а вырастает "снизу", из локальных взаимодействий.

Подход Хофстадтера немедленно подвергся критике ведущих представителей эвристической исследовательской программы – А. Ньюэлла и Э. Фейгенбаума. В искусственном интеллекте принято критиковать людей не за то, что они делают, а за то, чего они не делают. Поэтому Хофстадтера, ориентировавшегося на построение научной модели, бичевали с точки зрения инженерии – за отсутствие компьютерной программы, реализующей эту модель. "В искусственном интеллекте деньгами служат программы. Если вы не можете показать программу, то философы могут принять ваши идеи всерьез, но исследователи искусственного интеллекта – нет", – заявил Ньюэлл [Цит. по: Johnson,

1986, p. 290]. Хофстадтеру пришлось обороняться: "Сейчас я даже избегаю слов "искусственный интеллект". Я стараюсь называть свои исследования "когнитивной наукой". Искусственный интеллект теперь слишком загрязнен своими промышленными и коммерческими компонентами" [Цит. по: Johnson, 1986, p. 292]

Ранний коннекционизм, связанный с работами Ф. Розенблатта ("Персептрон") и О. Селфриджа ("Пандемониум"), базировался на нейрофизиологической модели человеческого мозга в виде случайной нейронной сети. Но устройства параллельного принципа действия, на которых был реализован персептрон, оказались вытеснены последовательными цифровыми компьютерами с "фон-неймановской" архитектурой, обладавшими более широкими вычислительными возможностями. Моделирование же параллельной обработки на цифровых ЭВМ было тогда чрезвычайно неэффективно.

В 1980-е годы развитие аппаратных средств совершенно изменило ситуацию. Во-первых, рост объема памяти и быстродействия "фон-неймановских" ЭВМ сделал возможным реальное моделирование параллельной обработки. Появилась возможность, "используя естественный параллелизм задачи, выразить его в явном виде, а затем с помощью компилятора перевести на любое имеющееся оборудование" [Фелдман, 1987, с. 147]. Во-вторых, были созданы эффективные компьютеры с параллельной архитектурой. В отличие от "фон-неймановской" ЭВМ, где центральный процессор отделен от блока памяти, в параллельных компьютерах одновременно работает от нескольких десятков до миллиона процессоров, обладающих собственной памятью. На такой аппаратуре реализуются модели "распределенной ассоциативной памяти", успешно решающие, например, задачу узнавания человека на фотографиях, сделанных с разных точек [Parallel..., 1981]. Эти работы в сообществе искусственного интеллекта принято объединять под названием "новый коннекционизм".

Притягательность коннекционистской исследовательской программы нашла выражение в притоке в ряды ее сторонников бывших адептов других программ. Исследователь из Стенфордского университета Д. Ленат с 1972 г. работал в группе К. Грина над системой автоматического программирования в рамках логистической исследовательской программы. Вскоре он начал разочаровываться в этом подходе. Логика, говорил Ленат, "делит мир на черное и белое, истину и ложь. Использовать логику – все равно, что использовать скальпель для разрезания мира. Вы не решаете уравнений гидродинамики, чтобы установить, выльется ли сейчас на вас вода. Вы используете некоторые грубые эвристики – ретроспективный взгляд на ваши жизненные неудачи. И вы каким-то образом собираете эту ретроспективу в небольшой набор правил, которые позволяют вам за доли секунды решить, нельзя ли избежать грядущей новой неудачи" [Цит. по: Johnson, 1986, p. 190].

Ленат обратился к эвристическому подходу, реализовав его в системе "Автоматизированный математик" ("Automated Mathematician" (AM)) [Lenat, 1982]. Система AM была предназначена для моделирования деятельности математика – введения новых математических понятий, выдвижения и проверки предположений, анализа примеров. Ленат использовал эвристики в виде правил-продукций "ЕСЛИ условие ТО действие", сформулированные для области теории множеств. Начав с небольшого набора исходных понятий, "Автоматизированный математик" самостоятельно "открыл" арифметику, понятие "простого числа", ряд известных теорем. Выяснилось, однако, что при этом начинает сказываться общий недостаток систем, основанных на правилах – резкое снижение эффективности работы при выходе за пределы исходной проблемной

области. Из первых 200 понятий, сгенерированных системой (в области теории множеств), 62,5% оказались нетривиальными, а из следующих 300 (в более широкой математической области) – лишь 10% [Ibid., p. 207]. Ленат вновь ищет выход и находит его в синергическом принципе мышления – "жестком ядре" коннекционизма.

В начале 80-х годов Ленат открывает возможность записать и понятия, и эвристики, к ним применяемые, в едином формате. Это позволяет применять любую эвристику к любой другой эвристике, как к понятию (грань между ними вовсе исчезает), и отказаться от жесткого разделения на данные и процедуры их обработки, характерного для всех булевских исследовательских программ искусственного интеллекта. Эта идея была воплощена Ленатом в системе EURISCO [Lenat, 1983], которая, как в свое время "Универсальный решатель задач", не была ориентирована на какую-то конкретную предметную область.

Для решения некоторой задачи EURISCO сообщаются лишь основные релевантные понятия-эвристики, далее работает механизм самообучения – применение собственных эвристик системы ко всей имеющейся информации, включая и сами эвристики. Оценка промежуточных результатов служит для увеличения ценности удачно проработавших эвристик и снижения ценности бесполезных и "вредных". EURISCO оправдала надежды, открыв ряд новых интересных понятий в области электроники и победив на двух национальных чемпионатах США по популярной интеллектуальной игре "Морская эскадра за два миллиарда" (в последующих чемпионатах компьютерам участвовать запретили). Результаты "Автоматизированного математика" EURISCO повторила, используя в 5 раз меньше эвристик и затратив вдвое меньше времени [Ibid., p. 95].

Источник силы своей программы Ленат видел в синергическом взаимодействии эвристик, превышающем эффект их простой суперпозиции [Lenat, 1982, p. 218]. Возможность для любой эвристики воздействовать на любую другую эвристику и, следовательно, на всю систему в целом, в том числе и вновь на себя саму, запускает типичный синергический механизм локальных взаимодействий, который, при отсутствии центрального управления (все эвристики равноправны), ведет к улучшению функционирования системы в целом. Особенность этих взаимодействий – в самообращенном характере связи между элементами системы (эвристиками), наличии внутренних обратных связей. Эту особенность можно проиллюстрировать на одном трагикомическом примере, когда EURISCO синтезировала новую эвристику, посоветовавшую уничтожить все эвристики, созданные системой. К счастью сама эта эвристика погибла одной из первых, так что вопрос решился сам собой.

Самообращенность внутренних взаимосвязей – когда связи между элементами системы образуют замкнутые цепочки – составляет суть небулевского интеллектуального механизма, реализуемого ныне коннекционистской исследовательской программой.

Если ранний коннекционизм опирался на нейрофизиологические модели, то текущие исследования идейно связаны с современной наукой о самоорганизации – синергетикой. Она изучает процессы возникновения пространственных, временных и пространственно-временных структур в нелинейных неравновесных системах. Источник нелинейности обнаруживается здесь также в самообращенности внутренних связей системы. Примером могут служить изучаемые школой И. Пригожина авто- и кросс-каталитические химические реакции. Автокаталитические петли, т. е. такие стадии, в которых продукт реакции участвует в синтезе самого себя, в уравнениях соответствуют нелинейным членам. В неравновесных условиях такая система способна рождать "порядок из хаоса"

[Пригожин и др., 1986].

Очень сходный механизм возникновения интеллекта предлагает упоминавшийся выше теоретик коннекционизма Д. Хофстадтер. Он пишет: "Ферменты распределяются случайно по цитоплазме клетки и постоянно наталкиваются на другие молекулы, собирают их и затем выполняют свои функции (синтезируют, перестраивают, разлагают)... Причем все это основано на случайном движении. Клетка находится на слишком низком биологическом уровне, чтобы здесь был некий интеллектуальный управляющий. 'Интеллект' клетки должен возникать из взаимодействия тысяч малых, независимых процессов. Волнообразная активность ферментов создает порядок из хаоса... Особенность клетки в том, что сами ферменты являются продуктами других ферментов..." [Hofstadter, 1984, p. 18]. В своей системе JUMBO, решающей анаграммы, Хофстадтер как раз применил "ферментную" процедуру случайного сочетания букв и синтеза слов.

Когнитивная исследовательская программа, заложившая основы когнитивной науки и ранее постоянно лидировавшая в научном споре программ (подчас уступая в инженерном), вынуждена была признать претензии коннекционизма. В лаборатории искусственного интеллекта МТИ развернулись работы над параллельными многопроцессорными компьютерами, а в 1983-84 гг. М. Минский пригласил поработать в МТИ Д. Хофстадтера. В последней книге Минского "Сообщество разума" центральный тезис уже взят из коннекционизма: "... разум можно построить из множества мелких частей, каждая из которых в отдельности неразумна... Каждый ментальный агент может выполнять лишь простую вещь, не требующую мышления. Но когда эти агенты особым образом объединяются в сообщество, это приводит к интеллекту" [Minsky, 1986, p. 17]. Минский указывает и на несводимый к простой суперпозиции (т.е. синергетический) характер взаимодействия этих ментальных агентов: "Жизнь, мышление – это отношение нескольких стенок, каждая из которых в отдельности не является "мышенепроницаемой", а ящик в целом является таковым — за счет особой организации стенок" [Ibid., p. 28].

В настоящее время три "классические" исследовательские программы искусственного интеллекта – эвристическая, логистическая и когнитивная – фактически слились в одну – "булевскую". Ей противостоит "небулевская" коннекционистская программа, которая явно прогрессирует по научным критериям оценки, пока еще уступая по инженерным.

## **Глава 3. РОЛЬ РЕФЛЕКСИИ УЧЕНЫХ В РАЗВИТИИ ИССЛЕДОВАНИЙ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ**

### **3. 1. Особенности рефлексии научного сообщества**

#### **3. 1. 1. Методологические проблемы анализа научной рефлексии**

Под рефлексией в науке принято понимать форму научного самосознания, направленную на анализ налично существующего знания и механизмов его формирования [Философский..., 1989, с. 556]. Каково же назначение такого анализа? М. А. Розов отмечает, что "рефлектирующие системы – это не просто системы, которые описывают сами себя. Их важной особенностью является то, что они способны управлять своим поведением в соответствии с полученным описанием, «способны как бы переключаться на новые механизмы управления" [Розов, 1987, с. 34]. Н. И. Кузнецова называет эту функцию рефлексии "самоконтролем", подразумевающим организацию и планирование поведения самой системы [Кузнецова, 1987, с. 214]. Кроме этого, Кузнецова выделяет еще одну функцию – указание на то, как следует воспринимать и анализировать эту систему, т.е. такое самоописание познавательных процедур, которое "подсказывает" наблюдателю их содержание: "Наука накапливает не только знание о природе, но и опыт познавательной деятельности. Без подключения к соответствующей рефлексии невозможно транслировать этот опыт..." [там же, с. 215].

Таким образом, рефлексия в науке выступает в двух ипостасях. Мы назовем их "рефлексией для себя" (для совершенствования собственной деятельности) и "рефлексией для других" (чтобы они могли правильно понять и воспроизвести мою деятельность). И если первая рефлексия, на наш взгляд, обращена на поиск нового и уделяет первостепенное внимание существующим недостаткам, то вторая скорее фиксирует положение дел и апеллирует к достоинствам. Заметим, что "рефлексия для других" вербализуется в научных публикациях, а "рефлексия для себя" чаще всего не выносится на всеобщее обозрение.

Описание систем научной деятельности с присущей им рефлексией требует применения специальной методологии, разработанной М. А. Розовым. Дело в том, что описание системы часто смешивают с ее самоописанием, с тем, какой она "хотела бы" выглядеть со стороны ("рефлексией для других"). В этом случае мы бы не анализировали систему, а лишь "переписывали" ее самооценку. Чтобы избежать этого, "исследователь должен занять ... особую надрефлексивную позицию, суть которой состоит в том, что сама рефлексия и ее функционирование делаются объектом изучения. Это значит, что мы должны описать некоторую деятельность до ее рефлексивного осознания, должны описать затем это последнее, должны показать, наконец, каким образом рефлексия приводит к перестройке исходной деятельности" [Розов, 1987, о. 37].

Пробуя применить эту методологию к анализу исследований по искусственному интеллекту, мы сталкиваемся с дополнительными проблемами, вызванными спецификой данной области. Мы уже отмечали сходство этих исследований с "неклассическими" научно-техническими дисциплинами (эргономикой, системотехникой). Представитель такой дисциплины, как отмечает В. Г. Горохов, разыгрывает сразу несколько профессиональных ролей: исследователя, инженера-проектировщика, методиста и т. д. "Эта 'многоликость' и в то же время единство профессиональных ролей приучает его

мышление к внутренней диалогичности и рефлексии, необходимости постоянно вставать в 'заимствованные позиции' участников кооперации, разрушает традиционную для классического естествознания и технической науки монологичность и монотеоретичность, стирает грани между исследованием и проектированием, собственно получением знаний и их использованием, между знанием и деятельностью" [Горохов, 1984, с. 67].

Ситуация, когда один и тот же человек одновременно выступает как ученый-психолог, лингвист, философ, инженер-программист, типична для области искусственного интеллекта. Можно сказать, что такой ученый как бы сам в себе содержит научное сообщество. Интересен новый, особый вид возникающей при этом рефлексии – назовем ее "мета-рефлексией". В данном случае деятельность существенно неоднородна: она включает элементы из различных научных традиций, и требуется особая работа рефлексии по их "увязыванию" в единое целое. Для такой "увязки" разнородных элементов, снабженных собственными рефлексивными описаниями, ученый вынужден подниматься на мета-уровень рефлексии и выступать как методолог (недаром столь велико число высказываний специалистов по искусственному интеллекту о взаимоотношениях психологии, логики, философии, лингвистики и других релевантных наук). Для "неклассических" научно-технических дисциплин, как указывает В. Г. Горохов, типична ситуация, когда профессиональная позиция ученого обязывает его "быть одновременно и методологом, постоянно осуществлять методологическую рефлексию собственной "синкретической" деятельности" [там же, с. 68].

Следовательно, при анализе исследований по искусственному интеллекту, где научное сообщество уже занимает "мета-рефлексивную" позицию, науковед должен придерживаться "над-мета-рефлексивной" позиции.

Казалось бы, ситуация уже достаточно сложна. Но следует учесть, что даже среди "неклассических" дисциплин область искусственного интеллекта выделяется своей оригинальностью. Ведь предмет изучения искусственного интеллекта как науки – само человеческое мышление. В известном смысле искусственный интеллект – это мышление о мышлении. Его модели можно интерпретировать как рефлексивные описания. В наиболее явной форме это выражено в моделях научной деятельности.

### **3. 1. 2. Моделирование научного познания**

Построением искусственного интеллекта занимались представители различных дисциплин – логики, психологии, лингвистики, биологии, заимствуя при этом модели соответствующих аспектов мышления. Но все они были прежде всего учеными, и им был одинаково хорошо знаком аспект мышления, связанный с научным познанием. Модель научного познания междисциплинарна, неспециализированна, общепонятна и выглядит для многих наиболее подходящей для реализации многофункционального искусственного интеллекта.

"Эвристический поиск" А. Ньюэлла и Г. Саймона [Ньюэлл и др., 1967в] был ничем иным, как компьютерным воплощением модели решения научных проблем, вербально сформулированной Дж. Пойа [Пойа, 1959, 1975].

"Эволюционное моделирование" Л. Фогеля [Фогель и др., 1969] опиралось на попперовскую индуктивистскую эпистемологию [Поппер, 1983]. Естественный отбор гипотез как метод научного мышления был обобщен при этом на всю деятельность интеллекта. Фогель писал: "По существу, механизация научного метода создает основу



для реализации искусственного интеллекта" [Фогель и др., 1969, с. 161].

Достижения философии науки 1960-х годов, оценившей роль научного сообщества в развитии исследований, также не остались незамеченными. М. Минский использовал куновскую концепцию "парадигмы" [Кун, 1975] как образец для формулировки понятия "фрейм" – стереотипной структуры для представления знаний [Минский, 1979]. Архитектура "доски объявлений" в системе понимания речи HEARSAY-II была основана на "метафоре научного сообщества": самостоятельные элементы системы ("источники знаний") кооперировались для решения общей задачи, публиковали свои результаты, использовали и критиковали чужие, экспериментировали, применяли знания и т. д. [Erman et al., 1980].

Э. Фейгенбаум также описывает свою экспертную систему DENDRAL как модель научного познания: "Блок-схема программы представляет собой замкнутый контур, в который включены блоки, представляющие этапы рассмотрения исходных данных, формирования гипотезы, прогнозирования и проверки гипотезы, т.е. она весьма точно соответствует контуру, характеризующему простейший 'научный подход'" [Фейгенбаум, 1973, с. 191-192].

Идеи "эволюционной эпистемологии" нашли развитие в системе EURISCO Д. Лената, открывавшей новые понятия в математике, теории игр и других областях [Lenat, 1983].

Хотя во всех перечисленных моделях научного познания имеющиеся образцы научной деятельности использовались как основа для оригинальных инженерных решений проблемы искусственного интеллекта, они же могут рассматриваться (и рассматривались) в качестве своеобразных теорий познания, записанных на машинном языке. Ленат, например, писал о "потенциальных возможностях EURISCO как модели познания" [Lenat et al., 1984, р. 292]. Можно сказать, что не только научная деятельность выступала в качестве образца при построении моделей мышления, но и системы искусственного интеллекта использовались как экспликации принципов научной деятельности, т.е. как рефлексивные описания.

Наиболее ярко это проявилось в случае с системой BACON Г. Саймона, П. Лэнгли и Г. Брэдшоу [Langley, 1979]. Пользуясь индуктивными правилами выведения количественных зависимостей между величинами, система "повторила" открытия Архимеда, Ома, Кеплера, Галилея. Для реализации также и дедуктивного вывода авторы заложили в систему законы сохранения массы, энергии, импульса. При этом преследуется не инженерная, а чисто научная цель – "изучение процессов научного открытия путем создания компьютерных программ, способных имитировать с большей или меньшей долей приближения тот путь, которым шел ученый" [Kalkarni, 1988, р. 139]. Для этого Г. Саймон с коллегами историко-научными методами восстанавливали все звенья цепочки рассуждений ученого и настраивали систему так, чтобы она точно воспроизводила их. В том же русле исследований Г. Саймоном и Д. Калкэрни была создана программа, моделирующая открытие орнитинового цикла Х. Кребсом (по историко-научной реконструкции Ф. Л. Холмса) [Ibidem]. Оборудованная соответствующими знаниями и эвристиками и помещенная в ту же когнитивную ситуацию, в которой находился Кребс, программа шаг за шагом повторила его путь.

В ответ на критику М. Сиксентмихальи [Csikszentmihalyi, 1988a, 1988b], обвинившего Г. Саймона в игнорировании основной составляющей творческого процесса – постановки проблем, последний вынужден был эксплицировать свои взгляды на научное

познание. Заметим, что в данном случае специалист по искусственному интеллекту для обоснования своего собственного научного подхода должен занять позицию ученого. Саймон высказал твердую установку на "демистификацию" творчества: "... и постановка проблемы, и поиск релевантных данных – все это процессы решения проблем нормального типа" [Simon, 1988, p. 178]. Многие ученые, в том числе и самые выдающиеся, утверждает он, отнюдь не изобретали тех задач, которые им удалось решить. В частности, применив к известной проблеме (синтеза мочевины) известный же метод (известный, однако, в другой области, ведь метод тканевых срезов применялся О. Варбургом для изучения клеточного дыхания), Х. Кребс открыл "орнитинный эффект" и объяснил синтез мочевины с помощью циклического ферментативного процесса — орнитинного цикла.

Модель научного познания, сформулированная Ньюэллом и Саймоном на заре эвристической исследовательской программы – это подбор подходящего метода к подходящей задаче (метод "анализа целей-средств"). Не так ли развивались и сами работы по искусственному интеллекту — двигаясь в "интеллектуальном пространстве" методов-задач? И разве случайно, что модель, выработанная в области искусственного интеллекта для общего описания научной деятельности, описывает и саму исследовательскую область? На наш взгляд, это вполне естественно, ибо научная рефлексия в данном случае не просто надстраивается над деятельностью ученых, но и входит в содержание этой деятельности.

### 3. 1. 3. Модели рефлексии

Коль скоро рефлексия является компонентой научного мышления в целом, то достаточно развитые системы искусственного интеллекта должны включать и модели рефлексии.

Когда к 1970-м годам системы искусственного интеллекта достигли высокого уровня сложности и стали содержать большое количество знаний, перед разработчиками встала проблема – как снабдить систему знанием о её знаниях. Человеку не надо перебирать весь свой интеллектуальный багаж, чтобы установить, знает ли он имя президента Гондураса. Для машины также необходим мета-уровень знаний – знания о собственных знаниях. В экспертных системах, построенных на основе продукционных баз знаний, роль мета-знаний выполняют мета-правила – правила применения правил. В эвристической исследовательской программе, развивавшей экспертные системы, была выработана, таким образом, иерархическая модель рефлексии.

В коннекционизме большее внимание было уделено самообращенному характеру рефлексии. Благодаря тому, что в программе Д. Лената EURISCO мета-правила записаны в том же формате, что и обычные правила, любая часть системы может "рефлектировать" над любой другой частью (в том числе и над собой). При этом возникает гетерархическая модель рефлексии. Теоретический анализ самообращенных явлений на материале логических построений К. Геделя, живописи М. Эшера и музыки И.-С. Баха предпринял идеолог коннекционистской исследовательской программы Д. Хофстадтер. В основе его модели интеллекта – рекурсивные построения, в которых каждая часть подобна целому, т.е. как бы отражает, "рефлектирует" целое. Особенностью интеллекта выступает включение самоописания в его собственное содержание. Ту же идею развивает Д. Дюбуа в своей "Фрактальной машине" – интеллектуальной компьютерной системе, базирующейся

на принципах фрактальной геометрии [Dubois, 1990].

В отечественной школе искусственного интеллекта одним из первых стал строить количественные модели рефлексии кибернетик М. Л. Цетлин [Поспелов Д. А., 1982; Цетлин, 1969]. В разработанной им схеме коллективного поведения автоматов было удобно изучать последствия "рефлексивного поведения" отдельных автоматов. В данной модели группа автоматов "питалась" из нескольких "кормушек", при этом каждый автомат, естественно, стремился получить максимум "пищи". "Рефлексивное поведение", по Цетлину, выражалось в том, что автомат сначала прогнозировал, какие действия совершат другие автоматы в коллективе, и лишь затем принимал решение о собственных действиях.

В модели Цетлина различались "ранги рефлексии". При нулевом ранге автомат выбирает действие лишь на основании сигналов от среды. При первом – предполагает, что остальные автоматы имеют нулевой ранг рефлексии, прогнозирует их поведение и на базе этого выбирает оптимальное действие для себя. При втором ранге рефлексии автомат предполагает наличие рефлексии первого ранга у остальных членов коллектива, опять прогнозирует и выбирает оптимум.

Введение рефлексии действительно привело к повышению разумности поведения. Автоматы с первым рангом рефлексии имели больший личный выигрыш, чем их коллеги с нулевым рангом, да и общий "улов" коллектива возрастал. Выявился при этом, однако, и обоюдоострый характер рефлексии. Излишняя рефлексия, как оказалось, приводила к потерям. Автоматы со вторым рангом рефлексии (приписывающие остальным первый ранг) в окружении "бездумных" автоматов (с нулевым рангом) вели себя менее эффективно, становясь жертвой самообмана. Снижался при этом и коллективный выигрыш.

Рефлексия коллектива автоматов в модели Цетлина имела гетерархическую структуру. Она складывалась из локальных взаимодействий индивидуальных рефлексивных актов отдельных автоматов. Основное отличие гетерархической модели рефлексии от иерархической – во включении в модель других рефлектирующих субъектов и образовании структур типа "Я думаю, что ты думаешь, что я думаю..." Когда отдельные элементы целого реализуют рефлексивные акты такого типа, их взаимодействие как раз рождает описанную выше "мета-рефлексию". Таким образом, особенности рефлексии научного сообщества искусственного интеллекта нашли отражение в конкретных моделях рефлексии, выработанных этим же сообществом.

### **3. 2. Виды деятельности и типы рефлексии в области искусственного интеллекта**

Итак, опишем вначале деятельность специалистов по искусственному интеллекту, затем их рефлексию и, наконец, единую рефлектирующую систему – научное сообщество, в котором изменения деятельности порождают перемены в самоописании, а изменения рефлексии трансформируют деятельность. Подытоживая вышеизложенный материал, можно выделить четыре вида деятельности ученых в области искусственного интеллекта:

1 – научное исследование и моделирование основных механизмов человеческого мышления в целом (это указывают своей целью М. Минский, А. Ньюэлл и Г. Саймон);

2 – научное исследование и моделирование человеческих методов решения отдельных задач (Т. Виноград, Д. Марр);

3 – научное исследование мощности логических исчислений (Дж. Маккарти);  
 4 – инженерное проектирование систем, решающих отдельные задачи (Э. Фейгенбаум, Э. Шортлиф).

Можно говорить о четырех "традициях работы", четырех способах ставить перед собой цель и извлекать результат. Но достаточно занять надрефлективную позицию, как разделение этих традиций становится условным. Конечный продукт каждой из них – компьютерная система, выполняющая некоторую интеллектуальную функцию. Рассматривая лишь сам этот продукт, можно ли с уверенностью сказать, по какой дороге он привезен, какая из традиций его породила?

Поставим этот вопрос перед конкретным ученым. Результатом его рефлексии может быть один из следующих ответов:

- 1 – эта система – модель человеческого мышления;
- 2 – модель человеческого метода решения отдельной задачи;
- 3 – исследование возможностей логико-математического формализма;
- 4 – компьютерная система прикладного назначения;
- 5 – работа в той предметной области, к которой относится данная интеллектуальная задача.

Получив такой ответ, сможем ли мы, однако, выбрать соответствующую традицию? Проблема в том, что разные ученые дадут различные ответы, да и один и тот же человек может иной раз иначе взглянуть на данную систему. Причин этого две:

1 – потенциальная полифункциональность каждой системы искусственного интеллекта;

2 – различия между типами рефлексии, дающие расхождения в оценке.

Мы уже рассматривали три типа рефлексии: "рефлексию для себя", "рефлексию для других" и "мета-рефлексию". Рефлективные описания часто соседствуют и сравниваются с оценками, которые дают внешние наблюдатели. Такие "сторонние" описания, необходимым образом присутствующие в обзорной и исторической литературе, в комментариях философов и психологов, включают (иногда косвенные) формулировки целей работы и критериев ее оценки, которые внешний наблюдатель дает как бы от имени автора.

Перечислим типичные противоречия, возникающие при наложении разных рефлективных и "сторонних" описаний:

1 – между "рефлексией для себя" и "рефлексией для других", когда определенные соображения заставляют автора представить свои результаты в иной форме (при этом могут преследоваться как чисто научные, так и сугубо конъюнктурные цели);

2 – между различными экспликациями "рефлексии для других" – при представлении работы в разных научных сообществах или для широкой общественности;

3 – между "рефлексией для себя" и "мета-рефлексией" – когда представления ученого о научном познании не согласуются с его собственной деятельностью; цикл рефлексии при этом не замкнут, и собственная деятельность не включается в рассмотрение при построении обобщающей модели научного познания;

4 – между "рефлексией для себя" и "сторонним" описанием – самое естественное противоречие, когда разные люди по-разному понимают одни и те же вещи.

Разберем несколько примеров "рефлективных противоречий" в исследованиях по искусственному интеллекту и проинтерпретируем их с точки зрения данной классификации.

Система "Логик-теоретик" (Logic Theorist) была создана А. Ньюэллом, Г. Саймоном и Дж. Шоу в 1955 г. Ее научной целью была разработка теории поискового решения задач, а инженерной – реализация машинного доказательства теорем исчисления высказываний в формализме Рассела-Уайтхеда. Сами авторы так формулировали свою задачу: "Цель настоящего исследования – разобраться в тех сложных процессах ("эвристиках"), которые участвуют в решении задач" [Ньюэлл и др., 19676, с. 113]. "Наши исследования машины "Логик-теоретик" являются звеном в общей программе исследования сложных систем переработки информации. Цель этой программы – разработка теории подобных систем и применение ее к таким областям, как составление машинных программ, обучение людей и решение задач" [там же, с. 114]. Итак, в "рефлексии для других" цель исследований формулируется без упоминания моделирования человеческого мышления, а инженерная цель выносится в область применений. Вся задача при этом как бы сводится к исследованию мощности соответствующих методов. Легко заметить, в то же время, что в данной работе сами "эвристики" являются моделями человеческих стратегий решения задач. Значит, в "рефлексии для себя" цель была иной – моделирование человеческих методов решения? Посмотрим теперь, по каким критериям авторы оценивают свою работу. К явным достижениям они относят число доказанных теорем (38), время доказательства (5-10 минут), эффективность пути доказательства (одна из теорем доказана более коротким путем, чем у Рассела и Уайтхеда). Не свидетельствует ли это о наличии в "рефлексии для себя" еще и инженерной цели?

Именно так рассудил американский математик Ван Хао, который отбросил "мишуру" искусственного интеллекта и расценил эту работу как чистый опыт в области доказательства теорем. При этом он предложил гораздо более эффективный формальный алгоритм доказательства (не используя никаких "эвристик") и добился замечательного инженерного успеха. На 52 теоремы у него ушло 5 минут машинного времени. Теоретическую нагруженность "Логика-теоретика" как модели человеческого мышления Ван Хао счел излишней, снижающей инженерную эффективность системы: "Нет необходимости стрелять из пушек по воробьям. Между тем, окончательное впечатление таково, что из своих пушек Ньюэлл, Шоу и Саймон не подстрелили даже воробья" [Ван Хао, 1962, с. 118]. В своем "стороннем" описании Ван Хао переинтерпретировал чужую работу в свете собственных целей и оценил ее по собственным критериям.

Описание, более адекватное авторской рефлексии, дал в данном случае М. Минский, защитивший работу Ньюэлла и Саймона от обвинений. Он заявил, что "некоторые из выпадов Ван Хао несправедливы. Его критика направлена не по адресу. Он, по-видимому, не понял, что авторов "Логика-теоретика" интересует не столько доказательство этих теорем, сколько общая проблема решения сложных задач... Задачи Ван Хао, будучи логически эквивалентными этим задачам, формально намного проще. Его методы не включают никаких способов использования ранее полученных результатов (следовательно, они неизбежно быстро потеряют силу при определенном уровне сложности проблемы), в то время как программа 'Логик-теоретик' ориентирована главным образом именно на это. Наконец, в результате того, что метод Ван Хао очень эффективен главным образом при решении специального множества теорем, ему просто не приходилось сталкиваться с основной эвристической проблемой: когда следует принять решение отказаться от того или иного пути продвижения к цели?" [Минский, 19676, с. 441-442].

Замечания Минского позволяют еще раз отметить важнейшую, на наш взгляд,

особенность исследований по искусственному интеллекту, корень многих сложностей их описания и самоописания. Работа по искусственному интеллекту не принадлежит целиком ни к моделированию человеческого мышления, ни к разработке эффективных методов решения задач. И та, и другая традиции работы потенциально присутствуют в исследовании, различаясь лишь отражением в рефлексии. В рефлексии всегда один элемент выступает как цель, а другой – как средство, одному придается первостепенное значение, другой же считается вспомогательным, один задаёт критерии оценки системы, другим при этом зачастую пренебрегают.

Так, например, постановка научной цели – создание теории мышления – предполагает использование определенных инженерных процедур проектирования, позволяющих разработать модель – компьютерную программу, воплощающую данную теорию. При этом можно требовать от этих процедур не эффективного решения задач, но лишь адекватности человеческим механизмам мышления. Э. Фейгенбаум и Дж. Фельдман так поясняют этот случай: "Исследователь..., заинтересованный в моделировании познавательной деятельности некоторого конкретного шахматиста во время игры, будет неудовлетворен, если его программа будет играть в шахматы лучше (или хуже) этого шахматиста, т. к. такой исследователь хотел бы, чтобы его программа делала бы те же ходы, что и данный шахматист, независимо от того, являются ли эти ходы хорошими или плохими" [Вычислительные..., 1967, с. 273]. Но в большинстве случаев от инженерных процедур все же требуют, чтобы система, моделируя человека, делала при этом хорошие ходы. Так, помимо научной, возникает (неявно) инженерная цель, что часто рождает противоречие.

С другой стороны, ставя чисто инженерную цель – создание эффективной интеллектуальной системы – ученый часто вынужден для этого моделировать человеческие методы решения задач, особенно когда вычислительные методы неэффективны или просто неприменимы. Дж. Слэйгл, например, считает, что "исследователь для построения успешно действующей машины вынужден стремиться к воспроизведению в машине методов, используемых людьми, т.к. люди обладают большим интеллектом, чем современные машины" [Слэйгл, 1973, с. 14]. Иногда (как в экспертных системах) приходится моделировать человеческое мышление, чтобы механика работы системы была понятна пользователю и допускала дальнейшее совершенствование. В данном случае уже научная цель выглядит вспомогательной в рефлексии автора такой системы. Посторонний наблюдатель, однако, может счесть ее и основной.

Каждый исследователь индивидуально выбирает, что считать основным, а что вспомогательным, но в любом случае ему приходится выполнять работу, варьируя на разных этапах критерии адекватности-эффективности. Поэтому так трудно указать единый критерий, поэтому постоянно возникают противоречия, поэтому исследования по искусственному интеллекту так легко защищать и так просто критиковать. И так трудно построить общую картину, в которую и критика, и защита укладывались бы как равноправные точки зрения.

Гетерогенность исследований по искусственному интеллекту зачастую делает их оценку с позиции любого фиксированного критерия одинаково негативной, независимо от того, какой критерий при этом выбирается. Сравним, например, два высказывания:

М. Рингль: "Хотя каждый из этих примеров нашел выражение в работающей компьютерной системе, все они являются скорее шагами к эксплицитной теории интеллекта, нежели вкладом в технику программирования" [Ringle, 1979, p. 11].

Дж. Вейценбаум: "Успехи 'искусственной интеллигенции' ограничиваются в основном триумфами технического характера. Их вклад в психологию познания или в практику решения задач невелик" [Вейценбаум, 1982, с. 293].

При этом оба диаметрально противоположных высказывания посвящены одним и тем же работам Т. Винограда, М. Минского и Р. Шенка! То, что в глазах одного наблюдателя выглядит успехом, для другого – неудача, и наоборот. Действительно, любая система искусственного интеллекта с точки зрения фиксированного критерия оценки может считаться неудачной, если обращать внимание на те стороны ее работы, которые не удовлетворяют этому критерию. Рассматриваемая же с другой стороны, она может выглядеть вполне успешной.

Что говорить о противоречиях рефлексивных описаний и сторонних оценок, где априорно трудно ожидать согласия, если даже сами создатели систем искусственного интеллекта иногда испытывают затруднения или впадают в противоречия при описании своих собственных работ. Э. Фейгенбаум, автор экспертной системы DENDRAL, задает сам себе риторический вопрос о ее научном статусе: "Относится ли программа, обеспечивающая выдвижение гипотез при анализе масс-спектров органических молекул, к искусственному интеллекту или к химии?" [Фейгенбаум, 1973, с. 173]. Он понимает, что его система включена одновременно в две традиции работы, и однозначный выбор между ними невозможен.

Показательно и противоречие в высказываниях М. Минского. С одной стороны, он отличает свой подход от попыток построения моделей поведения человека [Minsky, 19686, p. 7], а с другой – характеризует компьютерную программу своего ученика Т. Эванса "Аналогия", решающую геометрические задачи, как "теорию поведения человека в подобных ситуациях" [Ibid., p. 16]. Данное противоречие является всего лишь следствием того, что в первом случае М. Минский считал основной одну из целей своих исследований, а во втором – другую.

Анализируя рефлексивную научную общность искусственного интеллекта и обнаруживая в ней многочисленные формальные противоречия, мы склонны объяснять их не "ошибками" или "непониманием", а потенциальной полифункциональностью самого объекта рефлексии и сложностью и разнообразием рефлексивных процедур.

### **3. 3. О роли вопроса "Может ли машина мыслить?"**

- Может ли машина мыслить?
- Может.
- Так отчего же она не мыслит?
- А не хочет.
- Почему же не хочет?
- А она хотеть не может.

Фольклор "искусственной интеллигенции"

#### **3. 3. 1. Мировоззренческий смысл**

Трудно отыскать другой вопрос, занимающий в рефлексии специалистов по

искусственному интеллекту столь заметное и почетное место. Поставленный еще в статье А. Тьюринга 1950 г. [Тьюринг, 1960], он оживленно дебатруется до сих пор. Замечательно при этом то, что данная проблема не является чисто "внутренней" для области искусственного интеллекта, всем очевиден ее философский характер, но от искушения высказаться на эту тему не удержался ни один из крупных исследователей искусственного интеллекта.

Застарелая актуальность этого вопроса объясняется явной неудовлетворительностью как положительного, так и отрицательного ответов. Ответ "нет" обесмысливает саму идею искусственного интеллекта. Ответ "да" оказывается болезненным для человеческого самолюбия. Наука и так уже изрядно потрепала это самолюбие, нанеся, по выражению З. Фрейда, "три раны человеческому нарциссизму". Ими стали: гелиоцентрическая система Коперника (и современная космология в целом), теория Дарвина (и современная теория биологической эволюции) и, наконец, сам фрейдовский психоанализ. "В результате этих трех открытий выяснилось, что человек живет не в центре Вселенной, а на периферии одного из звездных скоплений; что он не создан по образу и подобию Божьему и, возможно, не уникален, а появился достаточно случайно и не так уж далеко ушел от животных; и наконец, что он не хозяин даже собственной психики, что ему не подвластны самые основы функционирования его сознания и, следовательно, его тела" [Шерток, 1990, с. 43]. Положительный ответ на вопрос "Может ли машина мыслить?" грозит нанести человеческому "нарциссизму" четвертую и, вероятно, самую глубокую рану, лишив человека монополии на разум. С одной стороны, это уже не позволяет человеку считать себя "венцом творения", а с другой – представляет совершенно практическую угрозу его господству.

Все это хорошо ощущают на себе исследователи искусственного интеллекта, сталкиваясь с контраргументами ненаучного характера или, как выразился Г. Саймон, "романтическим сопротивлением". "Доказывая, что машины мыслят, – писал он, — мы попадаем в такое же положение, что и Дарвин, когда он утверждал, что человек имеет общих предков с обезьянами, или Галилей, когда он доказывал, что Земля вращается вокруг своей оси" [Simon, 1990, p. 273].

Как любое крупное научное явление, исследования по искусственному интеллекту переходят рамки чисто познавательного феномена. "Я всегда так считал, и считаю сейчас, что наибольшее влияние искусственный интеллект оказывает на наше представление о себе", заявил Саймон в 1984 г. [Цит. по: Johnson, 1986, p. 56]. Правда, как и полагается адепту своей дисциплины, он истолковывает это влияние позитивно. Мы убеждаемся, утверждает Саймон, что являемся не чем-то особенным, а частью общего порядка вещей.

Последовательное проведение данной точки зрения приводит к представлению о человеке, как о машине, что обычно так возмущает противников искусственного интеллекта. Но отнюдь не смущает его защитников – достаточно познакомиться с аргументацией Н. Нильсена: "Гуманнее думать, что мы – машины. Разве человечно утверждать, что 'бедный Джонни не может выучить алгебру'? Если думать о Джонни как о машине, у которой ошибка в программе обучения алгебре, то это поможет ему гораздо больше. Мы не отправляем плохо работающий телевизор в тюрьму, а чиним его. Нет ничего страшнее, чем, имея неверную механическую модель человека, чинить его неправильно... Чем больше мы будем знать о себе как о механизмах, тем более человечными мы будем" [Цит. по: Johnson, 1986, p. 271-272].

Та же перспектива с иной точки зрения представляется гораздо менее радужной.



Известна притча про оптимиста и пессимиста, когда первый восклицает: "Это наилучший из всех возможных миров!", а второй вздыхает: "К сожалению, да". Дж. Вейценбаум перефразировал ее для искусственного интеллекта. Оптимист ликует: "Машины могут превзойти нас в умственном развитии!" Пессимист же в отчаянии повторяет ту же фразу [Вейценбаум, 1982, с. 266].

Возможность утратить интеллектуальное "первородство" для многих выглядит малопривлекательно. Исследования по искусственному интеллекту в таком случае представляются чем-то вроде рытья самим себе ямы или рубки собственного сука. Казалось бы, это вопрос философский, абстрактный, но следствия имеет самые практические. Несколько программ по созданию интеллектуальных роботов, начатых в конце 60-х годов, было заморожено, т.к. Управление перспективного планирования НИР Министерства обороны США прекратило их финансирование, как считают многие, из-за разговоров об опасности, которую представляют роботы [McCorduck, 1979, p. 233].

В результате вопрос "Может ли машина мыслить?" приобрел нравственный смысл. Дж. Вейценбаум выступил против идеи искусственного интеллекта (и компьютеризации в целом) не из-за ограниченности возможностей машин, а наоборот, из-за непомерной величины этих возможностей, грозящих, по его мнению, разрушить человеческую цивилизацию именно как "человеческую". Искусственный интеллект возможен, но безнравственен, – таково кредо Вейценбаума, прозванного за это "Торквемадой компьютеров". Нравственное переосмысление проблемы было четко сформулировано А. В. Шилейко: "Сейчас становится ясным, что основной вопрос состоит не в том, что могут и чего не могут вычислительные машины, а в том, что они должны и чего они не должны делать" [Дискуссии..., 1970, с. 46].

В то время, пока одни ставили философско-гносеологические (Х. Дрейфус), а другие – моральные (Дж. Вейценбаум) преграды искусственному интеллекту, сами исследователи искали практический ответ на вопрос.

### 3.3.2. Практический смысл

Ставя вопрос "Может ли машина мыслить?" как практический, исследователи искусственного интеллекта обязаны были учитывать гносеологические и моральные последствия своих действий. Поэтому они не только пытались позитивно решить его в философском плане, но и вести сами исследования так, чтобы избежать излишних нападков. Здесь мы сталкиваемся с ярким примером воздействия рефлексии на деятельность. Но если обычно эта "рефлексия для другого" бывает ориентирована на коллег, то в данном случае приходилось учитывать точку зрения внешних наблюдателей – философов, журналистов, широкой общественности.

Как отмечает Т. Кун, работа отдельного ученого обращена прежде всего к его коллегам, в то время как поэт или теолог обращается к непрофессиональной аудитории и ориентируется на ее оценки. "Именно потому, что он работает только для аудитории коллег – аудитории, которая разделяет его собственные оценки или убеждения, ученый может принимать без доказательств единую систему стандартов", т.е. парадигму [Кун, 1975, с. 207].

В исследованиях по искусственному интеллекту, напротив, необходимость учитывать философскую оценку и считаться с иными системами стандартов препятствует складыванию единой парадигмы. "Рефлексия для другого профессионального

сообщества" делает исследователя искусственного интеллекта ближе к поэту, чем к традиционному ученому.

Соотнесение исследований по искусственному интеллекту с философскими и психологическими системами стандартов проблематизирует употребление менталистских терминов типа "мышление", "понимание", "знание" в отношении машин. Здесь предлагается два выхода.

Одни, как М. Боден, предлагают использовать термины "зрение", "желание", "чувство" в метафорическом смысле: "Вычислительные теории разума – это лишь теории, и не более того. Никто не требует от химической теории, чтобы она растворялась в пробирке. Почему же нужно требовать, чтобы психологическая теория, заложенная в компьютер, видела и чувствовала?" [Boden, 1979, p. 130].

Другие исследователи предлагают отказаться от таких понятий, но применение взамен, скажем, физикалистских терминов (используемых нейрофизиологами) тут же породит такие же проблемы.

Наиболее радикальный выход, который нашло большинство исследователей искусственного интеллекта, – постулирование промежуточного между разумом и телом уровня описания. На таком уровне – информационном – вводятся "нейтральные" термины типа "информация", "программа", "структура данных", "правило". Однако, и для такой терминологии уже сложилась своя система стандартов в компьютерной науке, и перенести ее в область искусственного интеллекта непросто. "Чистые" программисты тоже имеют свой взгляд на искусственный интеллект как на пустую трату машинных ресурсов при реализации малоэффективных (с инженерной точки зрения) процедур.

В конце концов в области так и не сложилась устойчивая терминологическая система, так что результатом рефлексии явился не порядок, а "допарадигмальный" хаос.

Рефлексия, порождаемая вопросом "Может ли машина мыслить?", оказывает влияние на выбор не только терминологии, но и самих решаемых задач. На первом этапе развития области необходимо было продемонстрировать принципиальную возможность машинного мышления и выбирались наиболее выгодные задачи – доказательство теорем, игра в шахматы и т. п. Когда критики сочли эти задачи слишком простыми и алгоритмизируемыми, были намеренно избраны наиболее сложные и неформальные проблемы – машинное зрение, понимание естественного языка. Когда затем критики заявили, что эти задачи слишком туманны и четко не верифицируемы, была выбрана более строгая постановка – экспертиза в узкой профессиональной области и т. д. и т. п. По мере того, как "извне" уточнялось понятие "мышление", менялись требования к машинному мышлению и трансформировалась "внутренняя" ориентация исследователей.

Еще один аспект воздействия – критическая переоценка учеными используемых методов с точки зрения их соответствия представлениям об интеллектуальном рассуждении. Г. Геллернтер, автор нашумевшей программы доказательства геометрических теорем, прекратил дальнейшую работу над ней, считая, что программа "не отражает основных механизмов деятельности мозга" [Цит. по: Напалков, 1969, с. 65]. Его система доказывала теоремы, но не "мыслила".

Наконец, размышления над тем же "вечным вопросом" рождали и критерии оценки систем искусственного интеллекта. "Игра в имитацию", предложенная А. Тьюрингом в качестве строгой и корректной замены расплывчатого "Может ли машина..." [Тьюринг, 1960], была опробована как тест на системах ELIZA Дж. Вейценбаума, DOCTOR К. Колби и STUDENT Д. Боброва. Дж. Маккарти, в свою очередь, предложил заменить понятие

"мыслить" выражением "обладать здравым смыслом" (т. е. автоматически выводить следствия из уже известного) и в соответствии с таким критерием оценивал свою систему Advice Taker [McCarthy, 1968a].

Таким образом, деятельность и рефлексия постоянно корректировали друг друга: рефлексия задавала выбор терминологии, задач, методов и критериев оценки результатов, а деятельность требовала уточнения понятия "машинное мышление" и приводила ко все более строгой и определенной постановке "вечного вопроса".

### 3. 3. 3. Гносеологический смысл

Не случайно обсуждение вопроса "Может ли машина мыслить?" началось с его лингвистической критики и попытки переформулировать [Тьюринг, 1960]. Ведь он принадлежит к числу тех вопросов, ответы на которые заключены в прояснении, уточнении смысла формулировки самого вопроса. Действительно, проблема сводится к определениям понятий "машина", "мочь", "мыслить" и к их увязке. Все, однако, по-разному понимают эти слова, по-разному ставят вопрос и дают, естественно, разные ответы.

Каковы же типичные гносеологические проблемы, возникающие при постановке этого вопроса? Во-первых, трудно дать единое определение мышления, годящееся и для людей, и для машин. Как пишет Ю. М. Лотман, "мы имеем в своем распоряжении в качестве природного образца лишь один объект – мыслящее устройство человека. В силу своей единственности и несопоставимости объект этот создает для искусственного моделирования значительные трудности: почти невозможно решить, что в нем принадлежит Разуму вообще, а что следует отнести за счет случайностей его именно человеческой формы, возможно, весьма несовершенной" [Лотман, 1977, с. 12-13]. В теории рефлексии М. А. Розова эта проблема передается формулировкой "образец не задает множества реализаций" [Розов, 1987, с. 38]. Это означает, что отдельный образец (например, человеческий разум) не определяет сам по себе "множество похожих на него вещей", это зависит от выбора критерия сходства.

По каким же критериям сравнивать машинное поведение с человеческим, чтобы признать его разумным? Для Тьюринга таким критерием была имитация разговора, для Маккарти – выведение логических следствий, для Фейгенбаума – решение узкопрофессиональных задач. Интеллектуальные проблемы отличаются большим разнообразием, человеческие подходы к ним тоже различны; где же искать эталон? Как отметил А. В. Шилейко [Дискуссии..., 1970, с. 4 -5], мы рискуем при этом впасть в две крайности. Если относить к искусственному интеллекту любую машину, способную выполнить хоть одно интеллектуальное действие, то в эту категорию попадет и обыкновенный лифт. Если же отказывать в интеллектуальности машине, когда она не выполняет хотя бы одну функцию (будь то игра в шахматы или беседа об итальянском кинематографе), то такому жесткому критерию, скорее всего, не удовлетворит вообще ни один человек на Земле.

Известно, что тесты на коэффициент интеллектуальности оказываются неприменимы к людям разных культур – здесь нет единой точки отсчета. Поскольку и круг интеллектуальных задач, и стили мышления варьируются в разных культурах, то неясны рамки, ограничивающие сферу мышления как такового. Почему своеобразный разум вычислительных машин должен быть за пределами этой сферы, а непонятное нам

мышление дикарей из джунглей Амазонки – внутри нее?

Итак, с одной стороны, мы не хотим помыслить неантропоморфный разум, оберегая свой "нарциссизм", а с другой – и не можем этого сделать, ибо не знаем и, как показывает Лотман, не можем знать, где кончается человеческое мышление и где начинается Разум.

Вторая проблема, возникающая при постановке "вечного вопроса" – считать ли его теоретическим или эмпирическим, т. е. решать ли его априорно или апостериорно. Критики типа американского философа Х. Дрейфуса предпочитают решать его априори, говоря о принципиальной невозможности построения искусственного интеллекта [Дрейфус, 1978]. Специалисты по искусственному интеллекту, естественно, настаивают на эмпирическом решении. Еще в 1963 г. Э. Фейгенбаум и Дж. Фельдман писали:

"Может ли машина мыслить?"

Нет – если определять мышление как нечто, свойственное только и исключительно человеку.

Нет – если постулировать, что в самой сущности мышления есть что-то непостижимое, таинственное, мистическое.

Да – если признать, что этот вопрос должен быть решен путем эксперимента и наблюдения, путем сравнения поведения машин с поведением человека, с которым мы обычно связываем понятие 'мышление'" [Вычислительные..., 1967, с. 25].

На наш взгляд, чисто теоретическое негативное решение лишает вопрос смысла. Полагать априори, что не-человек не может мыслить, – значит счесть антропоморфность необходимым признаком интеллекта, не суметь отделить разум от его человеческой "оболочки". Какой смысл подходить к машине с критерием интеллектуальности, которому она в принципе не может удовлетворить только потому, что она не человек?

По нашему мнению, в вопросе "Может ли машина мыслить?" есть и теоретический, и эмпирический аспект, и отрыв их друг от друга рождает постоянное взаимное непонимание философов и исследователей искусственного интеллекта. Теоретический аспект находится в компетенции философов и состоит в выработке общего определения мышления и задании критериев оценки интеллектуальности. Это общее определение должно быть лишено антропоморфности и априорно допускать принципиальную возможность иного разума. Эмпирический аспект вопроса – дело специалистов по искусственному интеллекту – состоит в проверке, могут ли удовлетворить современные компьютерные системы данным критериям. До сих пор философы, как правило, формулировали критерии так, что им в принципе нельзя было удовлетворить, а исследователи искусственного интеллекта сами подстраивали критерии под возможности существующих систем, чтобы последние выглядели вполне интеллектуально.

Если теоретический аспект будет достаточно разработан философами и принят научным сообществом искусственного интеллекта, то дальше ответ уже зависит от прогресса исследований. Когда будет создана система, удовлетворяющая выработанным критериям, мы получим на "вечный вопрос" ответ "да". Но в любом случае мы никогда не услышим окончательного "нет".

## ЗАКЛЮЧЕНИЕ

Как справедливо заметили редакторы сборника "Вычислительные машины и мышление" Э. Фейгенбаум и Дж. Фельдман, "лучший способ изложения состоит в том, чтобы сначала рассказать аудитории, о чем вы собираетесь рассказывать, затем рассказать это, а в заключение еще раз сказать аудитории, о чем вы рассказали" [Вычислительные..., 1967, с. 401]. Попробуем кратко суммировать вышеизложенное.

Область искусственного интеллекта как объект науковедческого исследования обладает рядом специфических особенностей. На современном этапе развития исследования в этой области можно отнести к типу "неклассических" научно-технических дисциплин, в которых стирается грань между исследованием и проектированием. Система искусственного интеллекта – "двуликий" объект: в рефлексии ученых она может рассматриваться то как научное знание о механизмах человеческого мышления (или разума вообще), то как инженерный проект интеллектуальной машины. Поле таких исследований обладает сложной структурой, которую можно представить в виде двумерного пространства – проблемно-методической "решетки".

Традиционные логико-методологические схемы описания развития науки, предложенные Т. Куном и И. Лакатосом, в своем изначальном виде к истории искусственного интеллекта неприменимы. С точки зрения методологии парадигм, исследования в данной области все еще находятся на "допарадигмальной" стадии. Методология же исследовательских программ требует для своего применения существенной корректировки.

Мы предлагаем модифицированную схему конкуренции исследовательских программ в качестве модели развития области искусственного интеллекта. Роль "жесткого ядра" в таких программах играют исходные модели мышления, почерпнутые из смежных областей – психологии, логики, нейрофизиологии, биологии, эпистемологии. В качестве аналога сменяющих друг друга теорий выступают компьютерные системы, выполняющие все более тонкие интеллектуальные функции в рамках соответствующих исследовательских программ. "Позитивная эвристика" программы в данном случае состоит из двух частей: класса интеллектуальных проблем, определяемых областью, из которой взята исходная модель мышления, и набора программных средств, поставляемых соответствующей отрасли компьютерной науки.

Первая отличительная особенность модели связана с научно-инженерной двойственностью исследований, что делает неприменимыми однозначные (или научные, или инженерные) критерии прогресса или регресса программ. Вторая особенность заключается в разнообразии траекторий движения исследовательских программ по "интеллектуальному пространству" проблем-методов, что заставляет изменить понятие "конкуренции" программ.

Предлагаемый нами критерий прогресса программ предусматривает сравнение объема покрываемого разными программами "интеллектуального пространства". При этом предпочтение отдается исследовательским программам, выдвигающим новые проблемы и методы, т. е. расширяющим сами границы данного поля исследований. Оценки же по научным и инженерным координатам, на наш взгляд, должны производиться независимо, тем более, что они часто противоречат друг другу. Такая сложная система оценок соответствует реальной ситуации, когда в области искусственного интеллекта прогрессирует не одна, а несколько исследовательских

программ, при этом какая-либо программа, прогрессирующая с научной точки зрения, может вырождаться в инженерном плане, и наоборот.

Исследования по искусственному интеллекту зародились в лоне кибернетики, отделившись от нее в середине 1950-х годов. Точкой размежевания стал отказ от кибернетического структурно-функционального моделирования деятельности человеческого мозга и переход к чисто феноменологическому моделированию интеллектуальных функций, ориентированному на конечный результат деятельности. Важнейшей вехой в становлении исследований явился выбор компьютера в качестве средства реализации таких моделей. Возникло представление о едином классе систем переработки информации, куда входят и люди, и вычислительные машины.

В 1950-е гг. в области искусственного интеллекта возникло четыре исследовательские программы. Логистическая программа интерпретировала мышление как логический вывод в системе аксиом исчисления предикатов. Эвристическая – базировалась на психологической модели мышления как эвристического поиска в пространстве возможных состояний. Коннекционистская программа использовала нейрофизиологическую модель гетерархической организации нейронной сети, а эволюционная – биологические представления о случайных мутациях и естественном отборе. Начальный период развития исследований характеризовался широким разнообразием выбираемых интеллектуальных задач и отсутствием реальной конкуренции.

В 1960-е гг. "традиционные" программы испытали кризис при попытке распространить методы работавшие в искусственных, "лабораторных" задачах, на более широкий круг интеллектуальных проблем, возникающих в естественных человеческих ситуациях. В этих условиях прогресса добилась лишь новая, когнитивная исследовательская программа, основанная на эпистемологической исходной модели "мышление – это познание". Когнитивный подход был ориентирован на использование представлений человеческих знаний, позволяющих решать задачи машинного зрения и понимания естественного языка.

1970-е гг. ознаменовались жестокой борьбой исследовательских программ в рамках конкурса на создание лучшей компьютерной системы понимания речи. По инженерным критериям впереди оказалась эвристическая программа, но прогресса в научной постановке вопроса добилась скорее когнитивная. Расхождение научных и инженерных критериев оценки работ привело к разделению научного сообщества на школы "интеллектуальных артефактов" ("инженерии знаний") и "когнитивного моделирования". Этот раскол, на наш взгляд, произошел лишь в рефлексии научного сообщества, но не уровне конкретных исследований, которые сохранили свою потенциальную научно-инженерную двойственность.

С началом 1980-х гг., когда изделия искусственного интеллекта вышли на коммерческий рынок, в развитии эвристической, логистической и когнитивной программ стали преобладать инженерные цели; происходило техническое совершенствование систем при эксплуатации прежних теоретических схем. Существенного прогресса в научной постановке проблемы искусственного интеллекта добилась коннекционистская исследовательская программа, предложившая "небулевский" подход в противовес "булевскому" характеру остальных программ. Выработанная ею "синергическая" модель мышления позволила реализовать механизм самообучения и продуцирования разумного поведения за счет локальных взаимодействий неинтеллектуальных элементов.

Особое влияние на развитие исследований оказала рефлексия научного сообщества.

По самому статусу "неклассической" научно-технической дисциплины исследовательская деятельность в ее рамках должна включать рефлексивный момент. В области искусственного интеллекта, где предмет изучения – само мышление, рефлексия не просто надстраивается над деятельностью ученых, но и входит в ее содержание. Для исследователей из данной области их собственная работа служит источником моделей научного познания. В результате продукт исследований – компьютерная система – является как бы автотелью самих исследований.

Взаимодействие деятельности и рефлексии ученых в данной области осложняется большим разнообразием видов такой деятельности и типов рефлексии. По целевым установкам различаются следующие виды деятельности: моделирование основных механизмов мышления, моделирование человеческих методов решения отдельных задач, исследование мощности логико-математических формализмов и инженерное проектирование систем, решающих прикладные интеллектуальные задачи. Данное различие имеет смысл лишь в рефлексии ученых (где и формулируется цель исследования), реально же любой вид деятельности потенциально может быть переинтерпретирован с точки зрения иных целевых установок.

Разнообразие оценок деятельности порождается как самоописаниями (мы выделяем здесь три типа рефлексии: "рефлексия для себя", "рефлексия для других" и "мета-рефлексия"), так и "сторонними" описаниями. Сложность, двумерность "интеллектуального пространства" проблем-методов, полифункциональность исследований и, наконец, разнообразие рефлексивных процедур определяют поведение научного сообщества искусственного интеллекта как рефлектирующей системы.

Важнейшее место в рефлексии ученых занимает проблема "Может ли машина мыслить?" Ее нравственный смысл связан с угрозой утери человечеством своей монополии на разум. В то же время мировоззренческий аспект этой проблемы оказывает воздействие на чисто практическую сторону исследований. Необходимость учитывать философские и психологические критерии оценок заставляет ученых отказываться от традиционной терминологии, менять постановки задач, заново оценивать используемые методы, а на уровне "мета-рефлексии" вырабатывать новые, более ясные формулировки "вечного вопроса". Гносеологический смысл данного вопроса состоит в попытке выработать неантропоморфное, общее понятие "мышления". Проблема при этом распадается на теоретическую и эмпирическую части. Теоретическая задача заключается в формулировке общего определения мышления средствами философии. Эмпирическая часть состоит в проверке специалистами по искусственному интеллекту, способны ли компьютерные системы удовлетворить этому определению. Консенсус между философами и учеными до сих пор не достигнут.

Важнейшей нерешенной философской проблемой остается выработка такого общего, неантропоморфного определения мышления. Кроме того, философами еще недостаточно изучены мировоззренческие последствия исследований по искусственному интеллекту.

Что же касается науковедческого и историко-научного анализа данной области, то настоящая работа преследовала лишь цель задания каркаса – методологии и общей канвы дальнейших исследований. Подробной истории искусственного интеллекта еще предстоит быть написанной. До сих пор мало изучено влияние смежных дисциплин, но еще менее – обратное воздействие. Предварительным условием такого изучения, однако, была выработка адекватной модели развития самой области искусственного интеллекта как

целостного объекта. Мы попытались решить эту задачу, не скрывая ее сложности. Если нам это не удалось, то остается надеяться на времена, когда развитие рефлексивных способностей компьютеров позволит им самим написать диссертацию о себе.



## СПИСОК ЛИТЕРАТУРЫ

- Алексеева, 1987:** Алексеева И. Ю. Знание как объект компьютерного моделирования // Вопросы философии. – 1987. – №5. – с. 42-49.
- Арбиб, 1976:** Арбиб М. Метафорический мозг. – М.: Мир, 1976. – 295 с.
- Бенерджи, 1972:** Бенерджи Р. Теория решения задач. Подход к созданию искусственного интеллекта. – М.: Мир, 1972. – 224 с.
- Бирюков, 1978:** Бирюков Б. В. Что же могут вычислительные машины? Вместо послесловия // Дрейфус Х. Чего не могут вычислительные машины? – М.: Прогресс, 1978. – С. 298-332.
- Бирюков, 1985:** Бирюков Б. В. Жар холодных числ и пафос бесстрастной логики. формализация мышления от античных времен до эпохи кибернетики. – М.: Знание, 1985. – 192 с.
- Ван Хао, 1962:** Ван Хао. На пути к механической математике // Кибернетический сборник. – М.: ИЛ, 1962. – Вып. 5. – С. 2-22.
- Варшавский и др., 1984:** Варшавский В. И., Поспелов Д. А. Оркестр играет без дирижера. – М.: Наука, 1984. – 208 с.
- Вейценбаум, 1982:** Вейценбаум Дж. Возможности вычислительных машин и человеческий разум. От суждений к вычислениям. — М.: Радио и связь, 1982. – 367 с.
- Величковский и др., 1979:** Величковский Б. М., Зинченко В. П. Методологические проблемы современной когнитивной психологии // Вопросы философии. – 1979. – № 7. – С. 67-79.
- Величковский, 1982:** Величковский Е. М. Современная когнитивная психология. – М.: МГУ, 1982. – 336 с.
- Винер, 1968:** Винер Н. Кибернетика, или Управление и связь в животном и машине. 2-е изд. – М.: Советское радио, 1968. – 326 с.
- Виноград, 1976:** Виноград Т. Программа, понимающая естественный язык. – М.: Мир, 1976. – 294 с.
- Вудс, 1978:** Вудс В. Сетевые грамматики для анализа естественных языков // Кибернетический сборник. Новая серия. – М.: Мир, 1978. – Вып. 13.
- Вычислительные..., 1967:** Вычислительные машины и мышление Под ред. Э. Фейгенбаума и Дж. Фельдмана. – М.: Мир, 1967. – 552 с.
- Герович, 1989:** Герович В. А. Влияние идей философии науки на современные исследования по искусственному интеллекту // Формы представления знаний и творческое мышление. Тезисы Всесоюзного семинара. Часть 2. – Новосибирск, 1989. – С. 165-167.
- Герович, 1990а:** Герович В. А. Область искусственного интеллекта: науковедческий анализ // Синтез знаний: новый этап. Тезисы Всесоюзной научной конференции. – Одесса, 1990. – С. 45-46.
- Герович, 1990б:** Герович В. А. Проблема самоорганизации в кибернетике и искусственном интеллекте. — М.: ИИЕТ АН СССР, препринт № 32 (1990). – 36 с.
- Герович, 1990в:** Герович В. А. Развитие идеи самоорганизации в кибернетике и искусственном интеллекте // Материалы XXXII Всесоюзной научной конференции аспирантов и молодых специалистов по истории естествознания и техники. Часть 2. - М., 1990. – С. 266-269.
- Горелик и др., 1982:** Горелик А. Л., Гуревич И. Б. Предисловие к русскому изданию // Вейценбаум Дж. Возможности вычислительных машин и человеческий разум. От

суждений к вычислениям. М.: Радио и связь, 1982. – С. 3-21.

**Горохов, 1984:** Горохов В. Г. Методологический анализ научно-технических дисциплин. М.: Высшая школа, 1984. – 112 с.

**Гусев, 1986:** Гусев С. С. Активность человеческого сознания и "искусственный интеллект" // Вопросы философии. – 1986. – № 8. – С. 92-100.

**Дискуссии..., 1970:** Дискуссии об искусственном интеллекте / Автор-составитель А. В. Шилейко. – М.: Знание, 1970. – 48 с.

**Дрейфус, 1978:** Дрейфус Х. Чего не могут вычислительные машины? – М: Прогресс, 1978. – 336 с.

**Искусственный..., 1974:** Искусственный интеллект. Итоги и перспективы. – М., 1977-1987. – В 4 вып.

**Искусственный..., 1977-1987:** Искусственный интеллект. Аннотированный указатель литературы / Под ред. Л. Т. Кузина. – М., 1977-1987. – В 4 вып.

**Искусственный..., 1985:** Искусственный интеллект. Итоги и перспективы. – М.: МДНТП, 1985. – 154 с.

**Искусственный..., 1990:** Искусственный интеллект. Справочник / Под ред. Д. А. Поспелова и Э. В. Попова. – М.: Радио и связь, 1990. – В 3 кн.

**Кузин, 1974:** Кузин Л. Т. О системах искусственного интеллекта // Искусственный интеллект. Итоги и перспективы. – М.:МДНТП, 1974. – С. 3-15.

**Кузин, 1985:** Кузин Л. Т. Состояние и перспективы развития научно-технического направления "Искусственный интеллект" // Искусственный интеллект. Итоги и перспективы. – М.:МДНТП, 1985. – С. 4-19.

**Кузнецова, 1987:** Кузнецова Н. И. Научная рефлексия как объект историко-научного исследования // Проблемы рефлексии / Под ред. И. С. Ладенко. – Новосибирск: Наука, 1987. – С. 213-222.

**Кун, 1975:** Кун Т. Структура научных революций. – М.: Прогресс, 1975. – 288 с.

**Лакатос, 1978:** Лакатос И. История науки и ее рациональные реконструкции // Структура и развитие науки. Из Бостонских исследований по философии науки. – М.: Прогресс, 1978. – С. 203-269.

**Ленат, 1975:** Ленат Д. БИИНГИ: знание как система взаимодействующих экспертов // Труды IV Международной Объединенной конференции по искусственному интеллекту. – Тбилиси, 1975. – Т. 2. – С. 93-96.

**Ленат, 1984:** Ленат Д. Программное обеспечение искусственного интеллекта // В мире науки. – 1984. – № 11. – С. 112-122.

**Линдсей и др., 1974:** Линдсей П., Норман Д. Переработка информации у человека. – М.: Мир, 1974. – 550 с.

**Лотман, 1977:** Лотман Ю. М. Культура как коллективный интеллект и проблемы искусственного разума // Аннотации докладов Международной конференции по искусственному интеллекту. – Репино, 1977. – С. 12-14.

**Маккалок и др., 1956:** Маккалок У., Питтс У. Логическое исчисление идей, относящихся к нервной активности // Автоматы / Под ред. Дж. Маккарти и К. Шеннона. – М.: ИЛ, 1956. – С. 362--384.

**Маккарти и др., 1972:** Маккарти Дж., Хейес П. Некоторые философские проблемы в задаче построения искусственного интеллекта // Кибернетические проблемы бионики. – М.: Мир, 1972. – Вып. 2. – С. 40-88.

**Марр, 1987:** Марр Д. Зрение. Информационный подход к изучению представления и

обработки зрительных образов. – М.: Радио и связь, 1987. – 400 с.

**Минский, 1966:** Минский М. Проблемы в области искусственного интеллекта // Математические проблемы в биологии / Под ред. Р. Беллмана. – М.: Мир, 1966. – С. 78-98.

**Минский, 1967а:** Минский М. Библиография работ по искусственному разуму // Вычислительные машины и мышление / Под ред. Э. Фейгенбаума и Дж. Фельдмана. – М.: Мир, 1967. – С. 475-546.

**Минский, 1967б:** Минский М. На пути к созданию искусственного разума // Вычислительные машины и мышление / Под ред. Э. Фейгенбаума и Дж. Фельдмана. – М.: Мир, 1967. – С. 402-458.

**Минский, 1971:** Минский М. Вычисления и автоматы. – М.: Мир, 1971. – 364 с.

**Минский и др., 1971:** Минский М., Пейперт С. Перцептроны. – М.: Мир, 1971. – 261 с.

**Минский, 1979:** Минский М. Фреймы для представления знаний. – М.: Энергия, 1979. – 151 с.

**Найссер, 1981:** Найссер У. Познание и реальность: Смысл и принципы когнитивной психологии. – М.: Прогресс, 1981. – 230 с.

**Напалков, 1969:** Напалков А. В. Теория самоорганизующихся систем и эвристическое программирование // Бионические принципы самоорганизации. – Тбилиси: Мецниереба, 1969. – Вып. 1. – С. 65-72.

**Фон Нейман, 1960а:** фон Нейман Дж. Вычислительная машина и мозг

**Фон Нейман, 1960б:** фон Нейман Дж. Общая и логическая теория автоматов // Тьюринг А. Может ли машина мыслить? – М.: Физматгиз, 1960. – С. 59-101.

**Нильсон, 1973:** Нильсон Н. Искусственный интеллект. Методы поиска решений. – М.: Радио и связь, 1985. – 373 с.

**Ньюэлл и др., 1964:** Ньюэлл А., Шоу Дж., Саймон Г. Разновидности интеллектуального обучения GPS // Самоорганизующиеся системы / Под ред. М. Йовитса. – М.: Мир, 1964. – С. 211-257.

**Ньюэлл и др., 1967а:** Ньюэлл А., Шоу Дж., Саймон Г. Программа для игры в шахматы и проблема сложности // Вычислительные машины и мышление / Под ред. Э. Фейгенбаума и Дж. Фельдмана. – М.: Мир, 1967. – С. 33-71.

**Ньюэлл и др., 1967б:** Ньюэлл А., Шоу Дж., Саймон Г. Эмпирические исследования машины "Логик-теоретик"; пример изучения эвристики // Вычислительные машины и мышление / Под ред. Э. Фейгенбаума и Дж. Фельдмана. – М.: Мир, 1967. – С. 113-144.

**Ньюэлл и др., 1967в:** Ньюэлл А., Саймон Г. GPS – программа, моделирующая процесс человеческого мышления // Вычислительные мышление / Под ред. Э. Фейгенбаума и Дж. Фельдмана. – М.: Мир, 1967. – С. 283-301.

**Печенкин, 1984:** Печенкин А. А. Антифундаменталистские течения в западной "философии науки" // философские науки. Научные высшей школы. – 1984. – № 5. – С. 74-82.

**Пойа, 1959:** Пойа Д. Как решать задачу. – М.: Учпедгиз, 1959. – 207 с.

**Пойа, 1975:** Пойа Д. Математика и правдоподобные рассуждения. – М.: Наука, 1975. – 463 с.

**Поппер, 1983:** Поппер К. Логика и рост научного знания. – Прогресс, 1983. – 608 с.

**Поспелов Г. С. и др., 1974:** Поспелов Г. С., Поспелов Д. А. Искусственный интеллект: состояние и проблемы // Искусственный интеллект. Итоги и перспективы. М.: МДНТП, 1974. – С. 16--29

- Поспелов Г. С. и др., 1975:** Поспелов Г. С., Поспелов Д. А. Искусственный интеллект // Вестник АН СССР. – 1975. – № 10. – С. 26-36.
- Поспелов Г. С., 1980:** Поспелов Г. С. Возникновение и развитие искусственного интеллекта // Вопросы кибернетики. , 1980. – Вып. 61. – С. 5-12.
- Поспелов Д. А., 1982:** Поспелов Д. А. фантазия или наука: на пути к искусственному интеллекту. – М.: Наука, 1982. – 224 с.
- Поспелов Д. А., 1988:** Поспелов Д. А. Ближайшее будущее искусственного интеллекта // Тезисы Всесоюзной конференции по искусственному интеллекту. – Переславль-Залесский, 1988. – Т. 1. – С. 28-35.
- Построение..., 1987:** Построение экспертных систем / Под ред. Ф. Хейес-Рота, Д. Уотермана. – М.: Мир, 1987. – 441 с.
- Представление..., 1984:** Представление знаний в человеке-машинных и робототехнических системах. – М.: ВЦ АН СССР, 1984. – В 4 т.
- Пригожин и др., 1986:** Пригожин И., Стенгерс И Порядок из хаоса: новый диалог с человека с природой. – М.: Прогресс, 1986. – 432 с.
- Принципы..., 1966:** Принципы самоорганизации / Под ред. Г. фон Ферстера. – М.: Мир, 1966. – 622 с.
- Проблема..., 1987:** Проблема представления знаний системах (Материалы "Круглого стола") // Вопросы философии. – 1987. – № 1. – С. 52-61.
- Проблемы..., 1987:** Проблемы рефлексии / Под ред. И.С. Ладенко. – Новосибирск: Наука, 1987. – 238 с.
- Ракитов и др., 1986:** Ракитов А. И., Андрианова Т. В. Философия компьютерной революции // Вопросы философии. – №11. – С. 72-81.
- Ракитов, 1987:** Ракитов А. И. Методология историко-научных исследований в эпоху компьютерной революции // Вопросы истории естествознания и техники. – 1987. – № 3. – С. 32-41.
- Робинсон, 1970:** Робинсон Дж. Машинно-ориентированная логика, основанная на принципе резолюции // Кибернетический сборник. Новая серия. – М.: Мир, 1970. – Вып. 7. – С. 194-218.
- Розенблатт, 1964:** Розенблатт Ф. Обобщение восприятий по группам преобразований // Самоорганизующиеся системы / Под ред. М. Йовитса. – М.: Мир, 1964. – С. 65-106.
- Розенблатт, 1965:** Розенблатт Ф. Принципы нейродинамики. – М.: Мир, 1965. – 480 с.
- Розенблатт, 1966:** Розенблатт Ф. Стратегические подходы к исследованию моделей мозга // Принципы самоорганизации / Под ред. Г. фон Ферстера. – М.: Мир, 1966. – С. 469-482.
- Розов, 1977:** Розов М. А. Проблемы эмпирического анализа научных знаний. – Новосибирск, 1977. – 222 с.
- Розов, 1986:** Розов М. А. О двух аспектах проблемы редукционизма. – Пушкино: Научный центр биологических исследований, 1986. – 20 с.
- Розов, 1987:** Розов М. А. К методологии анализа рефлектирующих систем // Проблемы рефлексии / Под ред. И. С. Ладенко. Новосибирск: Наука, 1987. – С. 32-42.
- Рузавин, 1987:** Рузавин Г. И. Человек и робот // Вопросы философии. – 1987. – № 2. – С. 67-78.
- Саймон, 1972:** Саймон Г. Науки об искусственном. – М.: Мир, 1972. – 148 с.

- Самоорганизующиеся..., 1964:** Самоорганизующиеся системы / Под ред. М. Йовитса. – М.: Мир, 1964. – 436 с.
- Симонс, 1985:** Симонс Дж. ЭВМ пятого поколения: компьютеры 90-х годов. – М.: финансы и статистика, 1985. – 176 с.
- Слэйгл, 1973:** Слэйгл Дж. Искусственный интеллект. Подход на основе эвристического программирования. – М.: Мир, 1973. 319 с.
- Социально-философские..., 1979:** Социально-философские проблемы "человеко-машинных" систем. "Искусственный интеллект" как комплексная научно-техническая проблема (Материалы "Круглого стола") // Вопросы философии. – 1979. – №№ 2-4.
- Социальные..., 1986:** Социальные и методологические проблемы информатики, вычислительной техники и средств автоматизации (Материалы "Круглого стола") // Вопросы философии. -1986. – №№ 9-11.
- Тьюринг, 1960:** Тьюринг А. Может ли машина мыслить? – М.: Физматгиз, 1960. – 112 с.
- Уинстон, 1980:** Уинстон П. Искусственный интеллект. – М.: Мир, 1980. – 520 с.
- Фейгенбаум, 1973:** Фейгенбаум Э. Искусственный интеллект: темы исследований во втором десятилетии развития // Кибернетический сборник. Новая серия. – М.: Мир, 1973. – Вып. 10. – С. 171-203.
- Фелдман, 1987:** Фелдман Дж. Сетевые модели // Реальность и прогнозы искусственного интеллекта / Под ред. В. Л. Стефанюка. – М.: Мир, 1987. – С. 137-147.
- Философский энциклопедический словарь.** 2-е изд. – М.: Советская энциклопедия, 1989. – 815 с.
- Финн, 1983:** Финн В. К. О машинно-ориентированной формализации правдоподобных рассуждений в стиле Ф. Бэкона – Д. С. Милля // Семиотика и информатика. – М.: ВИНТИ, 1983. – Вып. 20. – С. 35-101.
- Фогель и др., 1969:** Фогель Л., Оуэнс А., Уолш М. Искусственный интеллект и эволюционное моделирование. – М.: Мир, 1969. – 230 с.
- Форсайт, 1987:** Форсайт Р. Феномен экспертных систем // Экспертные системы. Принципы работы и примеры / Под ред. Р. Форсайта. – М.: Радио и связь, 1987. – С. 9-14.
- Фэрли, 1964:** Фэрли Б. Самоорганизующиеся модели для обученного восприятия // Самоорганизующиеся системы / Под ред. М. Йовитса. – М.: Мир, 1964. – С. 19-46.
- Хант, 1978:** Хант Э. Искусственный интеллект. – М.: Мир, 1978. – 558 с.
- Хьюитт, 1987:** Хьюитт К. Открытые системы // Реальность и прогнозы искусственного интеллекта / Под ред. В. Л. Стефанюка. – М.: Мир, 1987. – С. 85-102.
- Цетлин, 1969:** Цетлин М. Л. Исследования по теории автоматов и моделированию биологических систем. – М.: Наука, 1969. – 316 с.
- Шалютин, 1985:** Шалютин С. М. Искусственный интеллект. – М.: Мысль, 1985. – 199 с.
- Шенк и др., 1975:** Шенк Р., Абельсон Р. Сценарии, планы и знание // Труды IV Международной Объединенной конференции по искусственному интеллекту. – Тбилиси, 1975. – Т. 6. – С. 208-220.
- Шенк и др., 1987:** Шенк Р., Хантер Л. Познать механизмы мышления // Реальность и прогнозы искусственного интеллекта / Под ред. В. Л. Стефанюка. – М.: Мир, 1987. – С. 15-26.
- Шерток, 1990:** Шерток Л. Гипноз – это толпа на двоих // Человек. – 1990. – № 1. – С. 42-49.

**Шиклоши, 1973:** Шиклоши Л. Об эволюции систем искусственного интеллекта // Кибернетический сборник. Новая серия. – М.: Мир, 1973. – Вып. 10. – С. 204-214.

**Шрейдер, 1990:** Шрейдер Ю. А. Интеллектуализация информационных систем // Итоги науки и техники. Серия "Информатика" / Под ред. А. И. Черного. – М.: ВИНТИ, 1990. – Т. 14. – С. 289-336.

**ЭВМ..., 1984:** ЭВМ пятого поколения. Концепции, проблемы, перспективы / Под ред. Т. Мото-ока. – М.: Финансы и статистика, 1984. – 110 с.

**Экспертные..., 1987:** Экспертные системы. Принципы работы и примеры / Под ред. Р. Форсайта. – М.: Радио и связь, 1987. – 224 с.

**Элти и др., 1987:** Элти Дж., Кумбс М. Экспертные системы: концепции и примеры. – М.: Финансы и статистика, 1987. – 191 с.

**Эндрю, 1985:** Эндрю А. Искусственный интеллект. – М.: Мир, 1985. – 264 с.

**Addanki, 1987:** Addanki S. Connectionism // Encyclopedia of Artificial Intelligence / Ed. by S. C. Shapiro. – Chichester: John Wiley, 1987. – P. 200-205.

**Advances..., 1988:** Advances in human-computer interaction / Ed. by H. Hartson and D. Dix. – Norwood: Ablex, 1988. – Vol. 2. – 380 p.

**Annals..., 1988:** Annals of the History of Computing. – 1988. – Vol. 10. – № 3.

**Arbib, 1987:** Arbib M. A View of brain theory // Self-organizing systems: The emergence of order / Ed. by F. E. Yates. – N.Y.: Plenum Press, 1987. – P. 279-311.

**Arnold et al., 1986:** Arnold W., Bowie J. Artificial Intelligence: A personal, commonsense journey. – Englewood Cliffs: Prentice-Hall, 1986. – 219 p.

**Artificial..., 1986:** Artificial Intelligence: methodology, systems, applications / Ed. by W. Bibel and B. Petkoff. – Amsterdam: North-Holland, 1986. – 247 p.

**Artificial..., 1987:** Artificial Intelligence: the case against // Ed. by R. Born. – L.: Croom Helm, 1987. – 220 p.

**Artificial..., 1988:** Artificial Intelligence and cognitive sciences / Ed. by J. Demongeot et al. – Manchester: Univ. Press, 1988. – 407 p.

**Banerji, 1980:** Banerji R. Artificial Intelligence: A theoretical approach. – Amsterdam: North-Holland, 1980. – 254 p.

**Bobrow, 1968:** Bobrow D. Natural language input for a computer problem-solving system // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 135-215.

**Boden, 1977:** Boden M. Artificial Intelligence and Natural Man. – N. Y.: Basic Books, 1977. – 537 p.

**Boden, 1979:** Boden M. The computational metaphor in psychology // Philosophical problems in psychology / Ed. by N. Bolton. – L.: Methuen, 1979. – P. 111-132.

**Bolter, 1984:** Bolter D. Turing's man: Western culture in the computer age. – Chapel Hill: Univ. of North Carolina Press, 1984. – 264 p.

**Bonnet, 1984:** Bonnet A. L'intelligence artificielle. – Paris: Inter. Ed., 1984. – 271 p.

**Bremermann, 1962:** Bremermann H. Optimization through evolution and recombination // Self-organizing systems / Ed. by M. Yovits et al. – Washington: Spartan Books, 1962. – P. 93-106.

**Breton, 1987:** Breton Ph. Histoire de l'informatique. – Paris: La Decouverte, 1987. – 239 p.

**Cognitive..., 1977:** Cognitive Science. – 1977. – Vol. 1. – № 1.

**Cohen et al., 1988:** Cohen P. R., Howe A. E. How evaluation guides Artificial Intelligence

research // Artificial Intelligence magazine. – 1988. – Vol. 9. – № 4. – P. 35-43.

**Courtial et al., 1989:** Courtial J. -P., Law J. A co-word study of Artificial Intelligence // Social studies of science. – 1989. – Vol. 19. – № 2. – P. 301-311.

**Csikszentmihalyi, 1988a:** Csikszentmihalyi M. Motivation and creativity // New ideas in psychology. – 1988. – Vol. 6. – № 2. – P. 159-176.

**Csikszentmihalyi, 1988b:** Csikszentmihalyi M. Solving a problem is not finding a new one: A reply to Simon // New ideas in psychology. – 1988. – Vol. 6. – № 2. . – P. 183-186.

**Davis et al., 1982:** Davis R., Lenat D. Knowledge-based systems in Artificial Intelligence. – N.Y.: McGraw-Hill, 1982. – 490 p.

**Dreyfus, 1979:** Dreyfus H. What computers can't do: The limits of Artificial Intelligence. Revised edition. – N.Y.: Harpers & Row, 1979. – 354 p.

**Dreyfus et al., 1986:** Dreyfus H., Dreyfus S. Mind over machine: e power of human intuition and expertise in the era of computer. – Oxford: Basil Blackwell, 1986. – 231 p.

**Dreyfus et al., 1988:** Dreyfus H., Dreyfus S. Making a mind versus modeling the brain: Artificial Intelligence back at a branchpoint // Daedalus. – 1988. – Vol. 117. – № 1. – P. 15-43.

**Dubois, 1990:** Dubois D. Le labyrinthe de l'intelligence: de l'intelligence naturelle a l'intelligence fractale. – Louvain-la-Neuve: Academia, 1990. – 322 p.

**Encyclopedia..., 1987:** Encyclopedia of the Artificial Intelligence / Ed. by S. C. Shapiro. – Chichester: John Wiley, 1987. – 2 vols.

**Erman et al., 1980:** Erman L. et al. The HEARSAY-II speech-understanding system: Integrating knowledge to resolve uncertainty // Computing Surveys. – 1980. – Vol. 12. – № 2. – P. 213-253.

**Ernst et al., 1969:** Ernst G., Newell A. GPS: A case study in generality and problem solving. – N.Y.: Academic press, 1969. – 297 p.

Evans, 1968: Evans T. A program for the solution of geometric-analogy intelligence test questions. // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 271-353.

Feigenbaum, 1971: Feigenbaum E. On generality and problem solving: a case study using the DENDRAL program // Machine Intelligence. – Edinburgh: Univ. Press, 1971. – Vol. 6. – P. 181-191.

**Feigenbaum et al., 1983:** Feigenbaum E., McCorduck P. The fifth generation: Artificial Intelligence and Japan's computer challenge to the world. – Reading: Addison-Wesley, 1983. – 275 p.

**Friedberg, 1958:** Friedberg R. A Learning Machine: Part I // IBM Journal of research and development. – 1958. – Vol. 2. - № 1. – P. 2-13.

**Friedberg et al., 1959:** Friedberg R. et al. A Learning Machine: Part II // IBM Journal of research and development. – 1959. – Vol. 3. – № 3. – P. 282-287.

**Gardner, 1985:** Gardner H. The mind's new science: A history of the cognitive revolution. – N.Y.: Basic Books, 1985. – 423 p.

**Garnham, 1988:** Garnham A. Artificial Intelligence. – London: Routledge and Kegan Paul, 1988. – 277 p.

**Handbook..., 1981-1982:** The Handbook of Artificial Intelligence / Ed. by A. Barr and E. Feigenbaum. – Los Altos: William Kaufmann, 1981-1982. – 3 vols.

**Haugeland, 1985:** Haugeland J. Artificial Intelligence: The very idea. – Cambridge, MA: MIT Press, 1985. – 287 p.

**Hayes-Roth, 1987:** Hayes-Roth B. Blackboard systems // Encyclopedia of Artificial

Intelligence / Ed. by S. C. Shapiro. – Chichester: John Wiley, 1987. – P. 73-80.

**Hofstadter, 1981:** Hofstadter D. Godel, Escher, Bach: An eternal golden braid. – Brighton: Harvester press, 1981. – 777 p.

**Hofstadter, 1984:** Hofstadter D. The Copycat Project: An experiment in nondeterminism and creative analogies. – Cambridge, MA: MIT AI Laboratory, 1984. – 35 p.

**Hofstadter, 1985:** Hofstadter D. Metamagical themas: Questing for the essence of mind and pattern. – N.Y.: Basic Books, 1985. – 852 p.

**Johnson, 1986:** Johnson J. Machinery of the mind: Inside the new science of Artificial Intelligence. – N.Y.: Times books, 1986. – 336 p.

**Kalkarni et al., 1988:** Kalkarni D., Simon H. The processes of scientific discovery: The strategy of experimentation // Cognitive science. – 1988. – Vol. 12. – № 2. – P. 139-175.

**Langley, 1979:** Langley P. Rediscovering physics with BACON // Proc. IJCAI 6. – Tokyo, 1979. – Vol. 1. – P. 505-507.

**Lauriere, 1987:** Lauriere J. L. L'intelligence artificielle. – Paris: Eyrolles, 1987. – 473 p.

**Lenat, 1982:** Lenat D. The nature of heuristics // Artificial Intelligence. – 1982. – Vol. 19. – № 2. – P. 189-249.

**Lenat, 1983:** Lenat D. EURISCO: A program that learns new heuristics and domain concepts // Artificial Intelligence. – 1983. – Vol. 21. – № 1-2. – P. 61-98.

**Lenat, 1984:** Lenat D. The role of heuristics in learning by discovery: three case studies // Machine learning: An Artificial Intelligence approach / Ed. by R. Michalski et al. – Berlin: Springer-Verlag, 1984. – P. 243-306.

**Lenat et al., 1984:** Lenat D., Brown J. Why AM and EURISCO appear to work // Artificial Intelligence. – 1984. – Vol. 23. – № 3. – P. 269-294.

**McCarthy, 1968a:** McCarthy J. Programs with common sense // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 403-409.

**McCarthy, 1968b:** McCarthy J. Situations, actions and causal laws // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 410-418.

**McCarthy, 1979:** McCarthy J. First order theories of individual concepts and propositions // Machine Intelligence. – Edinburgh: Univ. press, 1979. – Vol. 9. – P. 129-148.

**McCarthy, 1980:** McCarthy J. Circumscription – a form of non-monotonic reasoning // Artificial Intelligence. – 1980. – Vol. 13. – № 1-2. – P. 27-39.

**McCarthy, 1986:** McCarthy J. Applications of circumscription to formalizing common sense knowledge // Artificial Intelligence. – 1986. – Vol. 28. – № 1. – P. 89-116.

**McCarthy, 1988:** McCarthy J. Mathematical, logic in Artificial Intelligence // Daedalus. – 1988. – Vol. 117. – № 1. – P. 297-311.

**McCorduck, 1979:** McCorduck P. Machines who think: A personal inquiry into the history and prospects of Artificial Intelligence. – San Francisco: Freeman and Company, 1979. – 376 p.

**McCorduck, 1985:** McCorduck P. The universal machine: confessions of a technological optimist. – San Diego: Harcourt Brace Jovanovich, 1985. – 305 p.

**McCorduck, 1988:** McCorduck P. Artificial Intelligence: An aper-cu // Daedalus. – 1988. – Vol. 117. – № 1. – P. 65-83.

**Machine..., 1984:** Machine learning: An Artificial Intelligence approach / Ed. by R. Michalski et al. – Berlin: Springer-Verlag, 1984. – 572 p.

**Machine..., 1986:** Machine learning: An Artificial Intelligence approach / Ed. by R. Michalski et al. – Los Altos: Morgan Kaufmann, 1986. – Vol. 2. – 738 p.



**Marr, 1977:** Marr D. Artificial Intelligence – a personal view // Artificial Intelligence. – 1977. – Vol. 9. – № 1. – P. 37-48.

**Michalsky et al., 1984:** Michalsky R. et al. An overview of machine learning // Machine learning: An Artificial Intelligence approach / Ed. by R. Michalsky et al. – Berlin: Springer—Verlag, 1984. – P. 3-23.

**Mind's..., 1981:** The Mind's I: Fantasies and reflections on self and soul / Comp. and arranged by D. Hofstadter and D. Dennett. – N.Y.: Basic Books, 1981. – 501 p.

**Minsky, 1968a:** Minsky M. Descriptive languages and problem solving // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 419-424.

**Minsky, 1968b:** Minsky M. Introduction // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 1-32.

**Minsky, 1968c:** Minsky M. Matter, mind and models // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 425-432.

**Minsky, 1986:** Minsky M. The society of mind. – N.Y.: Simon and Schuster, 1986. – 339 p.

**Newell, 1962:** Newell A. Some problems of basic organization in problem-solving programs // Self-organizing systems / Ed. by M. Yovits et al. – Washington: Spartan Books, 1962. – P. 393-423.

**Newell et al., 1965:** Newell A., Simon H. Simulation of human processing of information. – Pittsburgh, PA: Carnegie Institute of Technology, 1965. – 8 p.

**Newell, 1970:** Newell A. Remarks on the relationship between Artificial Intelligence and cognitive psychology // Theoretical approaches to non-numerical problem solving / Ed. by R. Banerji and M. Mesarović. – Berlin: Springer-Verlag, 1970. – P. 363-400.

**Newell et al., 1972:** Newell A., Simon H. Human problem solving. – Englewood Cliffs: Prentice-Hall, 1972. – 920 p.

**Newell, 1973a:** Newell A. Production systems: models of control structures // Visual Information Processing / Ed. by W. Chase. – N.Y.: Academic Press, 1973. – P. 463-526.

**Newell, 1973b:** Newell A. You can't play 20 questions with nature and win // Visual Information Processing / Ed. by W. Chase. – N.Y.: Academic Press, 1973. – P. 283-308.

**Newell et al., 1973:** Newell A. et al. Speech understanding systems. Final report of a study group. – N.Y.: American Elsevier, 1973. – 137 p.

**Newell, 1982:** Newell A. The knowledge level // Artificial Intelligence. – 1982. – Vol. 18. – № 1. – P. 87-122.

**Newell, 1983:** Newell A. Intellectual issues in the history of Artificial Intelligence // The study of information / Ed. by F. Machlup and U. Mansfield. – N.Y.: Wiley, 1983. – P. 187-227.

**Nilsson, 1974:** Nilsson N. Artificial Intelligence // Proc. 4th IFIP Congress. – Stockholm, 1974. – P. 5-15.

**Nilsson, 1974:** Nilsson N. Artificial Intelligence: engineering, science or slogan? // Artificial Intelligence magazine. – 1982. - № 3. – P. 2-9.

**Papert, 1988:** Papert S. One artificial intelligence or many? // Daedalus. – 1988. – Vol. 117. – № 1. – P. 1-14.

**Parallel..., 1981:** Parallel models of associative memory / Ed. by R. Anderson and G. Hinton. – Hillsdale: Lawrence Erlbaum Assoc., 1981. – 295 p.

**Patent, 1986:** Patent D. The quest for artificial intelligence. – San Diego: Harcourt Brace Jovanovich, 1986. – 187 p.

**Philosophical..., 1979:** Philosophical perspectives in Artificial Intelligence / Ed. by M.

Ringle. – Brighton: Harvester press, 1979. – 244 p.

**Pratt, 1987:** Pratt V. Thinking machines: The evolution of Artificial Intelligence. – N.Y.: Basil Blackwell, 1987. – 254 p.

**Question..., 1988:** Question of Artificial Intelligence: Philosophical and sociological perspectives / Ed. by B. Bloomfield. – L.: John Wiley, 1988. – 432 p.

**Quillian, 1968:** Quillian M. Semantic memory // Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – P. 216-270.

**Rapaport, 1987:** Rapaport W. Logic // Encyclopedia of Artificial Intelligence / Ed. by S. C. Shapiro. – Chichester: John Wiley, 1987. – P. 536-538.

**Representation..., 1972:** Representation and meaning / Ed. by H. Simon and L. Siklossy. – Englewood Cliffs: Prentice-Hall, 1972. – 440 p.

**Rich, 1987:** Rich E. Artificial Intelligence // Encyclopedia of Artificial Intelligence / Ed. by S. C. Shapiro. – Chichester: John Wiley, 1987. – P. 9-16.

**Ringle, 1979:** Ringle M. Philosophy and Artificial Intelligence // Philosophical perspectives in Artificial Intelligence / Ed. by M. Ringle. – Brighton: Harvester press, 1979. – P. 1-20.

**Rose, 1984:** Rose F. Into the heart of the mind: An American quest for Artificial Intelligence. – N. Y.: Harper & Row, 1984. – 209 p.

**Rosenblatt, 1962:** Rosenblatt F. A comparison of several perceptron models // Self-organizing systems / Ed. by M. Yovits et al. – Washington: Spartan books, 1962. – P. 463-484.

**Schank, 1984:** Schank R. The cognitive computer: On language, learning and Artificial Intelligence. – Reading: Addison-Wesley, 1984. – 268 p.

**Scientific..., 1985:** The Scientific DataLink index to Intelligence research, 1954-1984. – N. Y.: Sci. Datalink, 1985. – 4 vols.

**Self-organizing..., 1962:** Self-organizing systems / Ed. by M. Yovits et al. – Washington, DC: Spartan Books, 1962. – 563 p.

**Self-organizing..., 1987:** Self-organizing systems: The emergence of order / Ed. by F. E. Yates. – N. Y.: Plenum Press, 1987. – 661 p.

**Selfridge, 1959:** Selfridge O. PANDEMONIUM: A paradigm for learning // Proc. Symp. on Mechanisation of Thought Processes. – Teddington, 1959. – P. 121-134.

**Semantic..., 1968:** Semantic Information Processing / Ed. by M. Minsky. – Cambridge, MA: MIT Press, 1968. – 438 p.

**Shortliffe, 1976:** Shortliffe E. Computer-based medical consultation: MYCIN. – N.Y.: American Elsevier, 1976. – 264 p.

**Simon, 1977:** Simon H. Models of discovery and other topics in the methods of science. – Dordrecht: Reidel, 1977. – 456 p.

**Simon, 1987:** Simon H. Models of man, social and rational: Mathematical essays on rational human behavior in a social setting. – N.Y.: Garland, 1987. – 287 p.

**Simon, 1988:** Simon H. Creativity and motivation: A response to Csikszentmihalyi // New ideas in psychology. – 1988. -Vol. 6. – № 2. – P. 177-181.

**Simon, 1990:** Simon H. Models of my life. – N.Y.: Basic books, 1990». – 316 p.

**Sloman, 1978:** Sloman A. The computer revolution in philosophy: philosophy, science and models of mind. – Sussex: Harvester Press, 1978. – 304 p.

**Special..., 1980:** Special issue on knowledge representation / Ed. by R. Brachman and B. Smith. – SIGART Newsletter. – 1980. – № 70.

**Stefik, 1981a:** Stefik M. Planning with constraints (MOLGEN: Part 1) // Artificial

Intelligence. – 1981. – Vol. 15. – № 1. – P. 111-139.

**Stefik, 1981b:** Stefik M. Planning and meta-planning (MOLGEN: Part 2) // Artificial Intelligence. – 1981. – Vol. 15. – № 2. – P. 141-169.

**Theoretical..., 1970:** Theoretical approaches to non-numerical problem solving / Ed. by R. Banerji and M. Mesarović. – Berlin: Springer-Verlag, 1970. – 468 p.

**Visual..., 1973:** Visual Information Processing / Ed. by W. Chase. – N.Y.: Academic Press, 1973. – 555 p.

**Winograd, 1983:** Winograd T. Language as a cognitive process. – Reading: Addison-Wesley, 1983. – Vol. 1: Syntax. – 640 p.

**Winograd et al., 1986:** Winograd T., Flores F. Understanding computers and cognition: A new foundation for design. – Norwood: Ablex, 1986. – 207 p.