

АКАДЕМИЯ НАУК СССР  
ИНСТИТУТ ФИЛОСОФИИ

На правах рукописи

ГЕРОВИЧ Вячеслав Александрович

ДИНАМИКА ИССЛЕДОВАТЕЛЬСКИХ ПРОГРАММ В ОБЛАСТИ  
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

специальность 09.00.09. - философские вопросы  
естествознания и техники

Автореферат диссертации на соискание ученой степени  
кандидата философских наук

---

Подписано в печать 21.11.91г. Зак. № 108-91  
Объем 1,3 уч. изд., 1,5 печ. л. Тир. 100 экз.  
Отпечатано на роталитге ИФАН, Волковка, 14.

Москва - 1991

Работа выполнена в Институте истории естествознания и техники АН СССР.

Научный руководитель: доктор философских наук А. А. Печенкин

Официальные оппоненты: доктор философских наук, кандидат физико-математических наук  
И. А. Шрейдер,  
кандидат философских наук  
В. И. Аршинов

Ведущая организация: кафедра философии и методологии науки  
Московского Государственного  
Университета (МГУ)

Защита состоится " \_\_\_\_ " \_\_\_\_\_ 1992 г. в 16 часов  
на заседании специализированного совета Д 002.29.03 по  
философским наукам в Институте философии АН СССР по адресу:  
Москва, ул. Волхонка, д. 14.

С диссертацией можно ознакомиться в библиотеке Института философии АН СССР.

Автореферат разослан " \_\_\_\_ " \_\_\_\_\_ 1991 г.

Ученый секретарь  
специализированного совета

Л. П. Князевко

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы определяется значимостью исследований по искусственному интеллекту, находящихся на переднем крае информатики и комплекса когнитивных наук. Проблема создания 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;
- историко-научное описание динамики области с использованием разработанной методологии;
- анализа 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, науковедческие модели и концепции, предложенные В. Г. Гороховым, Н. И. Кузнецовой, А. А. Печенкиным, А. И. Ракитовым, М. А. Розовым.

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 асимметрии предметного и рецензурного 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 научное сообщество. Поскольку сами исследования по искусственному интеллекту зародились в США, и американское 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: организации 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" [1990]. О публикациях 1970-х гг. дают представление выпуски "Аннотированного указателя литературы по искусственному интеллекту" [1977-1987]. Подробная библиография содержится в англоязычных справочных изданиях [Encyclopedia of Artificial Intelligence, 1987; Handbook of Artificial Intelligence, 1981-1982].

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 искусствен-

ного интеллекта посвящена не очень обширная литература, что выглядит вполне естественным, учитывая относительную молодость самой области. Непосредственно на эту тему написаны монографии Б. В. Вирюкова, Д. А. Поспелова, М. Боден, Дж. Хогеланда, Дж. Джонсона, П. Маккордак, В. Пратта и ряд статей Г. С. Поспелова, Э. Фейгенбаума, Л. Пиллоши, К. А. Шрейдера, П. Когена, А. Ньюэлла. В то же время довольно велик круг работ, где встречаются отдельные важные исторические и методологические замечания.

Часть историографии посвящена в основном предистории 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. Научометрические исследования, начатые в 1989 г. Ж.-П. Куртьелем и Дж. Лоу с применением метода совместной встречаемости слов в литературе по искусственному интеллекту, 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 подзадачам (общение на естественном языке, игры, распозна-

навание зрительных образов и т. д.) и последующее изложение каждого раздела как независимого направления исследований [М. Минский, Г. С. Поспелов, Э. Фейгенбаум]. 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 подзадачам" еще способно дать синхронический срез, "мгновенный портрет" дисциплины, то для отображения диахронического измерения, исторической динамики необходим иной подход.

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 рядом особенностей:

- результаты исследований по искусственному интеллекту двойственны: они могут рассматриваться и как научное знание о принципах человеческого мышления, и как инженерный проект интеллектуальной машины;

- предмет исследований имеет структуру двумерного пространства, образованного интеллектуальными проблемами и методами моделирования мышления.

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

3. Для анализа области необходима модификация методологии "исследовательских программ" И. Лакатоса, включающая уточнение понятий "жесткого ядра", "защитного пояса" и "позитивной эвристики" и переопределение критериев прогресса и регресса программ.

Роль "жесткого ядра" играет исходная модель мышления, черпнутая из смежной научной области, "защитного пояса" - ее конкретная реализация в виде компьютерной программы, а "позитивная эвристика" включает круг интеллектуальных проблем, характерных для той же смежной области, и набор релевантных программных средств.

4. Критерий прогресса исследовательских программ связан с объемом пространства проблем-методов, в котором разворачиваются программы. При этом регрессирует та исследовательская программа, что охватывает уже известные проблемы и методы, а прогрессирует программа, расширяющая само 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:

1950-е годы: возникновение первых четырех исследовательских программ, отделение от кибернетики.

1960-е годы: кризис "традиционных" исследовательских программ при переходе от лабораторных к широкому классу реальных задач. Возникновение когнитивной программы.

1970-е годы: Резкое расхождение научных и инженерных критериев оценки исследований. Если в инженерной области большого успеха добилась эвристическая программа, то в научной - когнитивная.

В 1980-е годы приверженцы логистической, эвристической и когнитивной программ производили инженерные изделия для коммерческого рынка, а по научным критериям прогрессировала коннекционистская программа.

7. Рефлексия специалиста по искусственному интеллекту входит в содержание его собственной деятельности: он анализирует свою работу, строит модели научного познания и рефлексии для интеллектуальных систем.

8. Противоречивая картина рефлексии научного сообщества искусственного интеллекта определяется следующими факторами:

- комплексной структурой предмета исследования (пространства проблем-методов);
- 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: влияет на 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 как методологическая основа для научно обоснованной оценки современного состояния и перспектив развития области искусственного интеллекта, для последующих исследований 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 использованной литературы. Список литературы включает 220 названий.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность диссертационной работы, определяются ее цель и задачи, раскрывается теоретико-методологическая основа исследования, его новизна и научно-практическая значимость, дается характеристика объекта изучения, хронологических рамок и источников, описывается структура работы и излагаются основные положения, выносимые на защиту.

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 М. Лакатосом, корректируется с учетом специфики исследуемой области.

Несмотря на то, что уже к началу 1970-х годов область искусственного интеллекта обрела все формальные признаки научной дисциплины (создание лабораторий, выпуск специализированных журналов, созыв конференций, чтение учебных курсов в университетах), для многих ведущих ученых вопрос о предмете и целях их исследований был и остается спорным. Выделяется пять различных точек зрения на характер исследований по искусственному интеллекту:

- 1) это набор разных научных и инженерных дисциплин (данное мнение разделяют Дж. Слейгд, Э. Рич и др.);
- 2) это инженерия (Н. Вильсон, Э. В. Попов);
- 3) это наука о человеческом мышлении (Р. Шенк, А. Ньюэлл, Г. Саймон), либо об общих принципах мышления, не обязательно связанных с человеческим объектом (М. Рингль);

4) это и наука, и инженерия одновременно (Р. Бенерджи, Дж. Джонсон);

5) это не наука и не инженерия, а нечто третье (скорее, искусство) (Х. Дрейфус, П. Уинстон).

Такое разнообразие мнений в рамках одного научного сообщества находит объяснение в сложности самого объекта - области искусственного интеллекта. Данная исследовательская область с науковедческой точки зрения обладает рядом следующих специфических особенностей:

- результаты исследований двойственны: они могут рассматриваться и как научное знание о механизмах мышления, и как инженерный проект интеллектуальной машины;
- предмет исследований имеет сложную структуру двумерного пространства - "решетки" 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.

Анализ уже имевших место попыток применить к описанию области логико-методологические схемы парадигм или научно-исследовательских программ показывает, что модели описания развития науки, предложенные Т. Куном и И. Лакатосом, в своем исходном 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 отраслью 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,двигающим новые проблемы и мо-

тоды, 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 стагнировать в инженерном плане, и наоборот.

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.

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 1960-е гг. в области искусственного интеллекта возникли четыре исследовательские программы. Логистическая программа (работы Дж. Маккарти) интерпретировала мышление как логический вывод в системе аксиом исчисления предикатов. Эвристическая

(А. Ньюэлл, Р. Саймон) - базировалась на психологической модели мышления как эвристического поиска в пространстве возможных состояний. Коннекционистская программа (Ф. Розенблатт, О. Селфридж, Б. Ферли, У. Кларк) использовала нейрофизиологическую модель гетерархической организации нейронной сети, а эволюционная (Р. Фридберг) - биологические представления о случайных мутациях и естественном отборе. Начальный период развития исследований характеризовался широким разнообразием решаемых интеллектуальных задач и отсутствием реальной конкуренции между программами.

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, развиваемый школой М. Минского, был ориентирован на использование представлений человеческих знаний, позволяющих решать задачи машинного зрения и понимания естественного языка.

1970-е гг. ознаменовались жесткой борьбой исследовательских программ в рамках конкурса на создание лучшей компьютерной системы понимания речи. По инженерным критериям впереди оказалась эвристическая программа (системы HEARSAY-II, HARPY Р. Редди), но прогресса в научной постановке вопроса добилась скорее когнитивная (система NWIM У. Вудса). 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, произошёл 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, логической 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 (работы Д. Хофстадтера, Д. Лената), предложившая "небулевский" подход в противовес "булевскому" характеру остальных программ, основанных на разложении интеллектуальной проблемы на подзадачи и их последовательном решении. Выработанная коннекционистской программой "синергическая" модель мышления позволила реализовать механизмы самообучения и продуцирования разумного поведения за счет локальных взаимодействий неинтеллектуальных элементов.

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.

Особое влияние на развитие исследований оказала рефлексия научного сообщества. По самому статусу "неклассической" науч-

но-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 автомоделью самих исследований.

Выработанные в области искусственного интеллекта модели мышления часто опирались на различные науковедческие схемы: эвристический метод решения задач Дж. Пола (работы А. Ньюэлла и Г. Саймона), индуктивистскую эпистемологию К. Поппера (подход Л. Фогеля), концепцию парадигм Т. Куна (работы М. Минского), эволюционную эпистемологию (подход Д. Лената). Данные модели можно рассматривать как своеобразные теории познания, записанные на машинном языке. По отношению к деятельности специалистов по искусственному интеллекту эти модели выступают в качестве рефлексивных описаний. Описаниями более высокого уровня являются созданные в данной области модели рефлексии (работы Д. Лената, Д. Дюбуа, М. Л. Цетлина). Они описывают действующего субъекта как рефлектирующую систему, вырабатывающую и активно использующую самописание.

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 мощностей 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 потери 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 этой проблемы. Гносеологический ее смысл состоит в попытке выработать неструктурное, общее понятие "мышления". Проблема при этом распадается на теоретическую и эмпирическую части. Теоретическая задача заключается в общем определении мышления средствами философии. Эмпирическая часть состоит в проверке специалистами по искусственному интеллекту, способны ли

компьютерные системы удовлетворить этому определению.

В заключении представлены основные выводы работы:

1. 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".

2. Развитие области искусственного интеллекта адекватно описывается модифицированной логико-методологической схемой исследовательских программ. Роль "жесткого ядра" в этих программах выполняют исходные модели мышления, взаимодействующие, соответственно, из логики (логистическая исследовательская программа), психологии (эвристическая), нейрофизиологии (коннекционистская), биологии (эволюционная) и эпистемологии (когнитивная).

Чисто научные или инженерные критерии прогресса программ часто противоречат друг другу; предлагаемый же нами интегральный критерий оценивает развитие программы по объему охваченной ею "решетки" решаемых интеллектуальных проблем и применяемых методов моделирования.

3. В 1960-е годы зародились первые исследовательские про-

граммы в области искусственного интеллекта - логистическая, эвристическая, коннекционистская и эволюционная. Выбрав для решения различные интеллектуальные проблемы и применяя разные методы моделирования, эти программы разошлись по "решетке" проблем-методов, не испытывая конкуренции.

В 1960-е годы эти программы замедлили развитие, столкнувшись с неадекватностью своих методов для решения новых задач. Прогресса добилась лишь новая, когнитивная программа.

В 1970-е годы по инженерным критериям прогрессировала эвристическая исследовательская программа, а по научным - когнитивная.

1980-е годы принесли научный успех коннекционистской программе, остальные же развивались в инженерном направлении, определяемом рынком сбыта изделий искусственного интеллекта.

4. Рефлексия научного сообщества искусственного интеллекта как "неклассической" научно-технической дисциплины включает в себя "мета-рефлексию" (увязку знаний и методов деятельности из различных смежных дисциплин) и сама, в свою очередь, включается непосредственно в деятельность ученого (при построении моделей научного познания, когда собственная деятельность выступает в качестве образца такого познания).

Постоянная дискуссия по проблеме "Может ли машина мыслить?" не является независимой "надстройкой" над конкретными исследованиями, а входит в качестве необходимого элемента в процесс постановки интеллектуальных задач, определения терминологии, уточнения критериев оценки систем. Рефлексия над этой проблемой является формой философского самосознания исследовательской области.

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. Б. М. Кедрова (Одесса, 1990), XXXII и XXXIII научных конференциях аспирантов и молодых специалистов по истории естествознания и техники (Москва, 1990, 1991). Работа обсуждалась на заседании сектора "История науки и логика" Института истории естествознания и техники АН СССР 4.6.91 г. и на заседании сектора теории познания Института философии АН СССР 25.6.91 г.

Основное содержание диссертационной работы изложено в следующих публикациях автора:

1. 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.

2. Область искусственного интеллекта: науковедческий анализ // Синтез знаний: новый этап. Тезисы Всесоюзной научной конференции. - Одесса, 1990. - С. 45-46.

3. 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.

4. Проблема самоорганизации в кибернетике и искусственном интеллекте. - М.: ИНИТ АН СССР, препринт N 62 (1990). - 66 с.